

Technical Report 547

19

AD A118253

**RESEARCH ISSUES IN THE DETERMINATION OF
SIMULATOR FIDELITY: PROCEEDINGS OF
THE ARI SPONSORED WORKSHOP
23-24 JULY, 1981**

Robert T. Hays (Editor)

SIMULATION SYSTEMS TECHNICAL AREA

DTIC
AUG 16 1982
H



U. S. Army

Research Institute for the Behavioral and Social Sciences

November 1981

Approved for public release; distribution unlimited.

82 08 16 032

DTIC FILE COPY

U. S. ARMY RESEARCH INSTITUTE FOR THE BEHAVIORAL AND SOCIAL SCIENCES

A Field Operating Agency under the Jurisdiction of the
Deputy Chief of Staff for Personnel

JOSEPH ZEIDNER
Technical Director

L. NEALE COSBY
Colonel, IN
Commander

NOTICES

DISTRIBUTION: Primary distribution of this report has been made by ARI. Please address correspondence concerning distribution of reports to: U.S. Army Research Institute for the Behavioral and Social Sciences, ATTN: PERI-TST, 5001 Eisenhower Avenue, Alexandria, Virginia 22333.

FINAL DISPOSITION: This report may be destroyed when it is no longer needed. Please do not return it to the U.S. Army Research Institute for the Behavioral and Social Sciences.

NOTE: The findings in this report are not to be construed as an official Department of the Army position, unless so designated by other authorized documents.

12

REPORT DOCUMENTATION PAGE

1. REPORT NUMBER Technical Report 547	2. GOVT ACCESSION NO. AD A118253	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) RESEARCH ISSUES IN THE DETERMINATION OF SIMULATOR FIDELITY: PROCEEDINGS OF THE ARI SPONSORED WORKSHOP 23-24 July, 1981	5. TYPE OF REPORT & PERIOD COVERED	
7. AUTHOR(s) Robert T. Hays (Editor)	6. PERFORMING ORG. REPORT NUMBER	
9. PERFORMING ORGANIZATION NAME AND ADDRESS US Army Research Institute for the Behavioral and Social Sciences (PERI-IE), 5001 Eisenhower Avenue, Alexandria, VA 22333	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 2Q162717A790	
11. CONTROLLING OFFICE NAME AND ADDRESS US Army Research Institute for the Behavioral and Social Sciences (PERI-IE), 5001 Eisenhower Avenue, Alexandria, VA 22333	12. REPORT DATE November 1981	
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)	13. NUMBER OF PAGES 156	
	15. SECURITY CLASS. (of this report) UNCLASSIFIED	
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE	
16. DISTRIBUTION STATEMENT (of this Report) Approved for public release; distribution unlimited		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		
18. SUPPLEMENTARY NOTES		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) Simulation training Simulators Simulator fidelity Fidelity requirements Fidelity research Fidelity guidance		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) This report is the proceedings from an ARI sponsored workshop on "Research Issues in Simulator Fidelity." Papers are included which: (1) state the goals and organization of the workshop, (2) present topics which were addressed during the workshop, (3) summarize the efforts of the working/discussion groups, and (4) summarize the results of the workshop as a whole. Workshop topics included definitional issues, communications issues, research support issues and research methodology issues.		

UNCLASSIFIED

1 SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

Technical Report 547

**RESEARCH ISSUES IN THE DETERMINATION OF
SIMULATOR FIDELITY: PROCEEDINGS OF
THE ARI SPONSORED WORKSHOP
23-24 JULY, 1981**

Robert T. Hays (Editor)

Angelo Mirabella
Team Chief

Submitted by:
Frank J. Harris, Chief
SIMULATION SYSTEMS TECHNICAL AREA

Approved by:
Milton S. Katz, Acting Director
TRAINING RESEARCH LABORATORY

U.S. ARMY RESEARCH INSTITUTE FOR THE BEHAVIORAL AND SOCIAL SCIENCES
5001 Eisenhower Avenue, Alexandria, Virginia 22333

Office, Deputy Chief of Staff for Personnel
Department of the Army

November 1981

Army Project Number
2Q162717A790

Human Performance Effectiveness and Simulation

Approved for public release; distribution unlimited.


ARI Research Reports and Technical Reports are intended for sponsors of R&D tasks and for other research and military agencies. Any findings ready for implementation at the time of publication are presented in the last part of the Brief. Upon completion of a major phase of the task, formal recommendations for official action normally are conveyed to appropriate military agencies by briefing or Disposition Form.

FOREWORD

The Simulation Systems Technical Area of the US Army Research Institute for the Behavioral and Social Sciences (ARI) performs research and development in areas that include training simulation with applicability to military training. Of special interest is research in the area of simulation fidelity requirements.

Before any training system may be developed and procured for use in the Army training community its specifications must be determined. These training device specifications, when compared to the actual equipment may be defined as simulator fidelity. It is necessary to determine the effects of level(s) of fidelity on training effectiveness if guidance is to be provided to support fidelity decisions.

This report documents a workshop, sponsored by ARI, which brought together individuals from the government, private industry, and academia to discuss "Research Issues in the Determination of Simulator Fidelity." The results of this report have implications for PM TRADE and for researchers in the areas of training and training device development.


JOSEPH ZEIDNER
Technical Director

Accession For	
NTIS GML1	<input checked="checked" type="checkbox"/>
DTIC TAB	<input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification	
By	
Distribution/	
Availability Codes	
Dist	Avail and/or Special
A	



RESEARCH ISSUES IN THE DETERMINATION OF SIMULATOR FIDELITY: PROCEEDINGS
OF THE ARI SPONSORED WORKSHOP 23-24 July, 1981

BRIEF

Requirement:

To document and disseminate the results of a two-day workshop sponsored by ARI on "Research Issues in the Determination of Simulator Fidelity."

Procedure:

Papers are included which:

- (1) State the goals and organization of the workshop.
- (2) Present topics which were addressed during the workshop.
- (3) Summarize the efforts of the working/discussion groups.
- (4) Summarize the results of the workshop as a whole.

Findings:

The question of the necessary level of training simulator fidelity cuts across many diverse areas of the Instructional Systems Development (ISD) process. Empirical research is needed to address the issues and answer the questions raised at the workshop. This research must systematically deal with all of the many variables which interact with training simulator fidelity to produce a given amount of transfer of training. Papers are included which state the issues and describe strategies for developing the necessary empirical data to deal with them.

Utilization of Findings:

This report can be used by researchers in determining how they may design empirical studies to determine necessary levels of training simulator fidelity and by the training community to aid in determining the nature of training devices, which must be designed, procured, and implemented.

RESEARCH ISSUES IN THE DETERMINATION OF SIMULATOR FIDELITY: PROCEEDINGS
OF THE ARI SPONSORED WORKSHOP 23-24 JULY, 1981

CONTENTS

	Page
Goals and Organization of the Training Simulator Fidelity Workshop Robert T. Hays.....	1
ARI Simulation Research Program Angelo Mirabella.....	9
SIMTRAIN Contract Overview Stephen F. Hirshfeld.....	21
A Framework and Topics for Empirical Research on Training Simulator Effectiveness David R. Baum.....	25
Problems in Simulation for Training Robert T. Hennessy.....	39
Summary of Major Ideas from Thorpe's Presentation "Constraints on Fidelity" Robert T. Hays.....	57
A Fidelity Definition and Outline of Interactive Factors Robert T. Hays.....	65
FIDELITY: What We Know and Who Should Know It Clarence A. Semple.....	77
Effectiveness Issues Rohn J. Hritz.....	83
Definitions in Simulation Worth Scanlan.....	89
Is Simulation Fidelity the Question? William E. Montague.....	93

Simulator Use in Four Functional Areas	
Leon H. Nawrocki.....	111
Summaries of Topic Areas in the Working Group Sessions	
Effectiveness Session	
Garvin L. Holman.....	121
Fidelity Session	
Thomas J. Houston.....	125
Guidance Session	
Gary A. Klein.....	129
Research Priorities and Support Session	
Robert W. Swezy.....	135
Research on the Fidelity of Simulators	
Jesse Orlansky.....	143
Summary Remarks	
David R. Baum.....	153
The ARI Fidelity Workshop: Summary Comments	
Robert T. Hays.....	159
List of Attendees.....	167

GOALS AND ORGANIZATION OF THE
ARI TRAINING SIMULATOR FIDELITY WORKSHOP

Robert T. Hays
U.S. Army Research Institute

A workshop on training simulator fidelity was held on the 23rd and 24th of July, 1981, at the Army Research Institute for the Behavioral and Social Sciences, in Alexandria, Virginia. The workshop was sponsored and organized by ARI as part of its research program on guidelines for the design of training simulators. ARI was assisted in this effort by Honeywell, Science Applications Inc., and Klein & Associates. In attendance were 67 personnel (see list of attendees), including uniformed representatives of the Army, Navy, and Air Force, and civilian representatives from industry, government and academia.

The goal of the workshop was to define and prioritize research issues for a program of empirical research on the relationship between simulator fidelity and training effectiveness. This goal was based on the premise that increasing numbers of training needs can be met more effectively through the use of operational equipment or "platform" techniques. But, to maximize the cost-effectiveness of simulators, it is necessary to define, through hard, scientific data, the lowest amount of fidelity relative to actual equipment which will result in satisfactory levels of training.

The workshop goal was pursued through invited presentations followed by a series of working groups on focused topics. This strategy was designed to encourage a free exchange of information, opinions, and perspectives on research issues relevant to training simulator fidelity. It was also designed to encourage practical planning for empirical research, rather than theoretical discussion. Each group was given the task of answering two major questions in a topic area relevant to training simulator fidelity. Table 1 lists the four topic areas and the questions that were addressed by the working groups. To insure that all participants had the opportunity to express their views on all topics, a rotational strategy was followed which cycled everyone through all four groups. The first working session was twice as long as subsequent sessions since participants in later sessions could utilize the work of their predecessors. The products of the working groups were summarized at the end of the workshop for the group as a whole.

The workshop proceeded in four phases. Table 2 provides the workshop agenda and also shows the four phases of the workshop. The first phase consisted of short presentations which located the workshop in the larger context of the Army's, ARI's, and the Simulation Systems Technical Area's program of research on training simulation. In this phase the main goals of the workshop were stated to the participants. It was stated that the purpose of the workshop was to brainstorm about how to conduct useful research on training simulator fidelity and that the emphasis of the workshop was on work.

The second phase of the workshop consisted of a series of short presentations which only raised issues because the strategy was to hold tentative answers for discussion in the working

groups. The purpose of this phase was to sensitize everyone to certain major issues and also to clarify the topics which would be addressed later.

The third phase was the working group meetings themselves. Each working group had a topic leader and an archivist. Their job was to insure that the group addressed the stated questions in a given topic area and to provide a summary of the group's discussions during the final phase of the workshop. In addition to the topic leader and archivist, who remained with a given topic area, there was a group leader who rotated with a number of participants through all four topic areas and insured that the interpersonal functioning of the working groups was facilitated.

In the final phase, summaries of the working group discussions were presented. The last phase also included two overall summaries of the workshop.

Included in the remainder of this document are the orientation presentations, the research issue presentations, the topic leader summaries, and the workshop overviews. In a final summation paper, conclusions were drawn about the workshop itself, as well as about the possible directions of future research in the area of training simulator fidelity.

TABLE 1

Working Group Topic Areas and Discussion Questions

Group I: Effectiveness Topics

- A: Transfer Effectiveness vs. Other Measures of Effectiveness
 - (1) What are the types of effectiveness measures?
 - (2) Under which conditions is one measure chosen over another?
 - (3) What are the criteria for establishing effectiveness?
- B: Cost Effectiveness
 - (1) What is cost effectiveness?
 - (2) How do we measure it? Is there a cost effectiveness metric?
 - (3) How does cost effectiveness relate to training?

Group II: Fidelity Topics

- A: The Measurement of Fidelity
 - (1) What is Fidelity?
 - (2) It is measureable? If so, what kind of metric? Why measure it?
 - (3) What is the relationship, if any, between fidelity and training?
- B: Generalizability of Flight Simulation Data to Other Areas of Simulation Training

Given that most data on the relationship between training simulator fidelity and effective training has come from the realm of flight simulation:

 - (1) Can we generalize flight simulation data to other areas of simulation training?
 - (2) What are the components/elements/factors that influence the generalizability of these data?
 - (3) What are the conditions under which generalizability is possible?
 - (4) How do we use the flight simulation data that is generalizable?

Group III: Guidance Topics

- A: Development of a Format for Fidelity Decision Making Guidance

Given that the persons who must specify the characteristics to be incorporated into a training device need guidance in making fidelity decisions:

 - (1) On what issues/factors do they need this guidance?
 - (2) Is guidance available?
 - (3) In what format should this guidance be provided?
- B: Impact of New Technologies on Fidelity Decisions

Given that new technologies are being developed which could affect training strategies:

 - (1) How well do we use new technologies?
 - (2) How can we better incorporate new technologies into training systems?
 - (3) Is it possible/desirable for training devices to keep us with new technologies?

Table 1 (continued)

Group IV: Priorities and Support Topics

A: Ranking Fidelity Research Issues in Terms of Necessary Resources vs. Payoffs.

Given that there are innumerable research issues relating to training simulator fidelity:

- (1) How do we determine which issues to address and the order in which to address them?
- (2) Do different groups rank issues differently? In what ways do they differ? Why?

B: Methods for Generating Long-Term Research Support and Communication

- (1) How do we gain necessary support to undertake a long-term research effort to answer the questions raised in this workshop.
- (2) How can we facilitate the communication of data among all interested parties from the user level through DoD?

TABLE 2
WORKSHOP AGENDA

"RESEARCH ISSUES IN THE DETERMINATION OF
SIMULATOR FIDELITY"

23-24 July 1981

23 July

Phase I	0745	Coffee and Registration	
	0830	Call to Order	Dr. Robert T Hays, ARI
	0835	ARI Welcome	COL Franklin Hart, Commander ARI
			Dr. Joseph Zeidner, Technical Director, ARI
	0840	ARI Simulation Research Program	Dr. Angelo Mirabella, ARI
	0850	SIMTRAIN	Dr. Stephen Hirshfeld, Honeywell
	0900	Fidelity Determination - Task Objectives and Structure	Dr. David Baum, Honeywell, SRC
	0915	Workshop Objectives and Organization	Dr. Robert T. Hays, ARI
			John Brock, Honeywell, SRC
	0925	Issues in Simulation	Dr. Robert Hennessy, National Research Council
	0940	Constraints on Fidelity	MAJ Jack Thorpe, DARPA
	0955	A Fidelity Definition and Interactive Factors	Dr. Robert T. Hays, ARI
	1010	BREAK	
Phase II	1030	What we know and who should know it	Clarence Semple, Canyon Research
	1045	Effectiveness Issues	Dr. Rohn Hritz, Applied Science Assoc.
	1100	Definitions in Simulation	Dr. Worth Scanland, CNET (Navy)
	1115	Is Simulation Fidelity the Question?	Dr. William Montague, NPRDC
	1130	Simulator use in 4 Functional Areas	Dr. Leon Nawrocki, ARI
	1145	CLOSE OF PRESENTATION SESSION	
	1200	LUNCH	
	1300	Working Group Organization and Assignments	John Brock, Honeywell ARC
	1315	Working Group Session I	
	1600	BREAK	
	1615	Working Group Session II	
	1745	END OF FIRST DAY	

Phase III

24 July

	0800	Coffee and Rolls	
	0830	Working Group Session III	
	1000	BREAK	
	1015	Working Group Session IV	
	1145	LUNCH	
Phase IV	1245	Informal Information Exchange/ Topic Leaders prepare Summations	
	1345	Topic Leaders present Summations	
	1500	BREAK	
	1515	Closing Observations	Dr. Jesse Orlansky, IDA
	1545	Closing Remarks	Dr. David Baum, Honeywell SRC
			Dr. Robert T. Hays, ARI
			John Brock, Honeywell SRC

ARI SIMULATION RESEARCH PROGRAM

Angelo Mirabella
U.S. Army Research Institute

RESEARCH PAGE BLANK-NOT FILMED

ABSTRACT

The Army Research Institute workshop on fidelity is located in the larger context of its research program in training simulation. The objectives of the Simulation Systems Technical Area of ARI are explained with focus on the three major tasks which are relevant to the workshop.

Good morning. I would like to add my own note of welcome to those sounded by COL Hart and Dr. Zeidner. The workshop agenda states that I will talk about the ARI Simulation Research Program. But I could not do justice to such a complex topic in the few minutes allotted to me. I would like to explain how this workshop on fidelity fits into a much larger research program at ARI. That program is organized under five major headings, one of which is labeled Training For Combat Effectiveness (TFCE). In turn, TFCE is organized into smaller divisions, one of which is called the Simulation and Training Devices Thrust (SATD). Research on Simulator Fidelity represents one small, but, we think, significant work unit within the SATD Thrust.

The objectives of the SATD Thrust include those listed in Table 1. Cost-effectiveness in the design of simulators and training devices and improved methods for evaluating simulator/training device effectiveness are major goals. More specific objectives are listed in the lower portion of Table 1.

Though Fidelity is only one of many Training System issues which must be addressed in achieving these objectives, we feel its role is central and critical, if only because it has been a common meeting ground for engineers, psychologists, and training specialists.

ARI is organized to carry out simulation systems research as shown in Figure 1. That organization includes a headquarters element, the Simulation Systems Technical Area, which is sponsoring this workshop. It also includes ARI Field Units and Forts Benning, Knox, Bliss, Sill, Rucker and Leavenworth. The targets of this research include air, air defense, armor, infantry, artillery, and maintenance training.

But at this point I would like to focus on the objectives of the Simulation Systems Technical Area since the program of that particular unit has inspired the organization of this workshop. There are three tasks under this unit which are of particular relevance (Table 2): one task is aimed at constructing methods and models for developing training device requirements and evaluation device effectiveness. A second task is concerned with methods and models for defining simulation fidelity requirements. And, a third has to do with providing support to PM TRADE, in evaluating the Army's AMTESS*.

The objective of the first task (Table 3) is to develop methods and models for both prescribing device requirements and for predicting device effectiveness. The primary approach has

*Army Maintenance Training and Evaluation Simulation System

TABLE 1
TRAINING FOR COMBAT EFFECTIVENESS
SIMULATION AND TRAINING DEVICES THRUST

OBJECTIVES

- **COST-EFFECTIVE SIMULATORS AND TRAINING DEVICES**
- **EVALUATION OF TRAINING DEVICES THROUGHOUT LIFE CYCLE OF TRAINING SYSTEMS**
- **STRATEGIES FOR INTEGRATING LIVE FIRE AND SIMULATION**
- **REDUCTION IN TIME, FUEL, PHYSICAL SPACE**
- **GENERALIZED SIMULATION SYSTEMS**
- **REALISTIC TRAINING VIA ADVANCED TECHNOLOGY**

FIGURE 1
ORGANIZATION OF SIMULATION AND TRAINING DEVICES PROGRAM

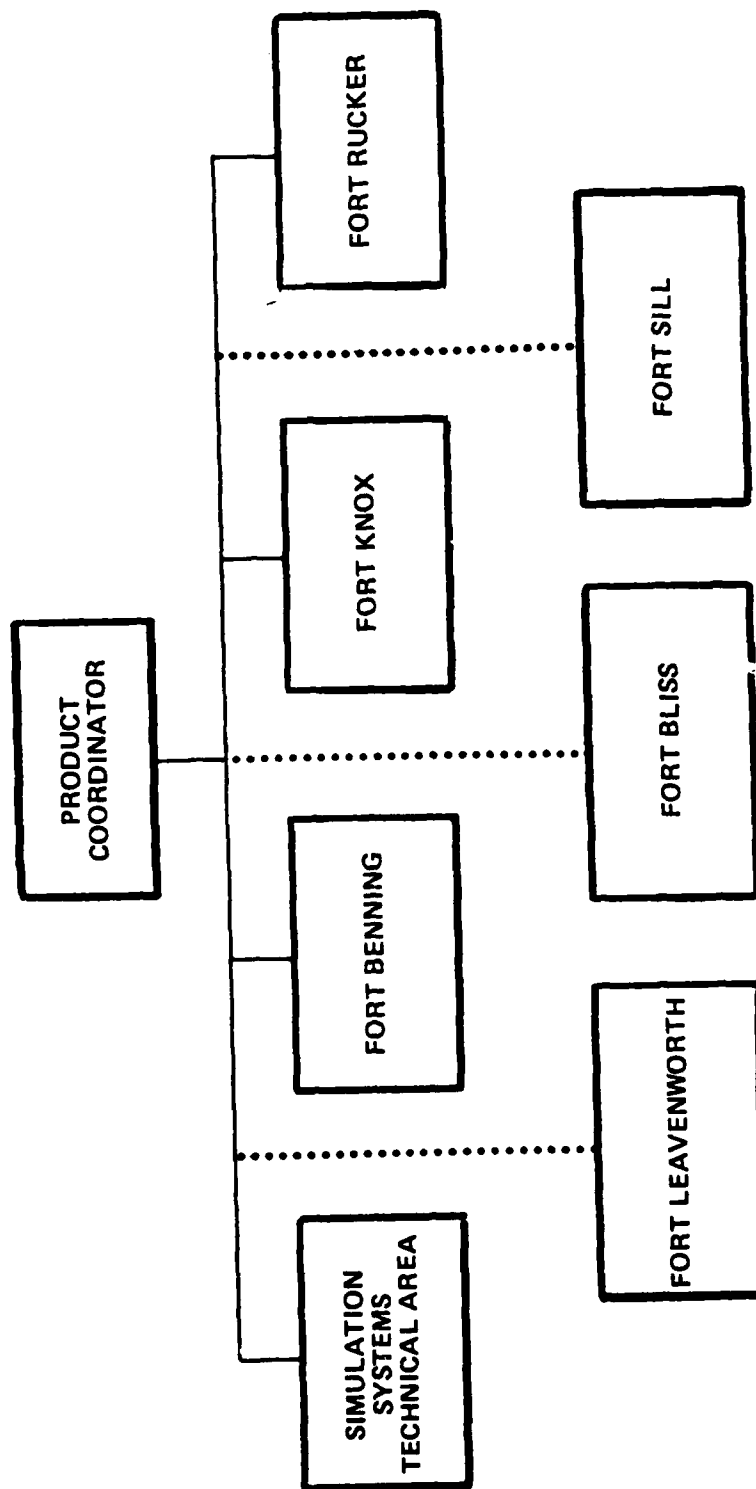


TABLE 2

MAJOR SSTA RESEARCH TASKS

- DEVELOPMENT OF TRAINING DEVICE REQUIREMENTS -- METHODS AND MODELS.
- DEFINING SIMULATION FIDELITY REQUIREMENTS -- METHODS AND MODELS.
- ASSESSING AND UTILIZING ARMY MAINTENANCE TRAINING AND EVALUATION SIMULATION SYSTEM (AMTESS).

TABLE 3

TASK 1: DEVELOPMENT OF TRAINING DEVICE REQUIREMENTS

OBJECTIVE:

MODELS AND METHODS FOR PRESCRIBING/PREDICTING
DEVICE REQUIREMENTS/EFFECTIVENESS

APPROACH:

ANALYZE/TEST/DEVELOP EXISTING MODELS, e.g., TRAINVICE
I & II.

- FEASIBILITY
- RELIABILITY
- VALIDITY

PRODUCTS:

- PROCEDURES FOR ESTIMATING DEVICE EFFECTIVENESS.
- PROCEDURES FOR SPECIFYING TRAINING DEVICE REQUIREMENTS.

been to further analyze, test, and develop existing models such as TRAINVICE I and II with respect to the feasibility of their being used in the field, their reliability and their validity. TRAINVICE, a model developed by ARI, yields a figure of merit to describe the effectiveness of a training device (Wheaton, et al, 1976). The anticipated products include procedures for estimating device effectiveness and procedures for specifying training device requirements.

The second task (Table 4), which perhaps is most pertinent to the work that we will be doing over the next two days, is titled "Defining Simulation Fidelity Requirements". Its objective is to develop methods and models for conducting fidelity requirements analyses, through a four-step approach. The first step is to explore existing literature and data for assistance in constructing useful guidelines. This particular step is unlikely to be very productive for reasons documented by Waag (1981). Waag's conclusion is that the parametrically rich laboratory studies needed to develop comprehensive and detailed guidelines have not been conducted in spite of a 30-year history of fidelity research. Such studies are envisioned as step 2. A long-range outcome of such studies could be a computerized data base on the relationships between fidelity and training effectiveness. Related outcomes would be a model and procedures for conducting a fidelity requirement analysis. The end product envisioned is a set of user-oriented guidelines for generating fidelity requirements. The last of the three particularly relevant tasks (Table 5) supports PM TRADE in evaluating alternative AMTESS prototypes. Empirical tests of the training effectiveness of the prototypes will be conducted along with an analysis of particular pieces of architecture of those prototypes. User problems attitudes vis-a-vis the employment of those prototypes will also be explored. The end product will include a data base and recommendations for modifying, producing, and using AMTESS.

These three tasks have, to a large extent, inspired and given impetus both to this workshop and to the various issues that we are going to be covering over the next two days. I must confess that we have a somewhat selfish objective in putting the workshop together. We would like to pick the brains of people who know at least as much as we do, in many cases more than we do, so that we can do a better job of planning and conducting our research under these three tasks. However, the results that emerge from this conference will be community property. We certainly hope that everyone will leave the conference with some better insights, some better information on the issues to be discussed, information which will help you in thinking about your own research or device acquisition responsibilities. Once again, we thank all of you for joining with us in these deliberations.

TABLE 4

TASK 2: DEFINING SIMULATION FIDELITY REQUIREMENTS

OBJECTIVE:

DEVELOP MODELS AND METHODS FOR CONDUCTING FIDELITY REQUIREMENTS ANALYSIS.

APPROACH:

- ABSTRACT PRINCIPLES FROM EXISTING STUDIES.
- CONDUCT LABORATORY STUDIES.
- DEVELOP COMPUTERIZED DATA BASE ON FIDELITY
- FORMULATE A MODEL AND PROCEDURE FOR FIDELITY ANALYSIS.

END PRODUCT:

USER ORIENTED GUIDELINES FOR GENERATING FIDELITY REQUIREMENTS

TABLE 5

TASK 3: ASSESSING AND UTILIZING ARMY MAINTENANCE
AND EVALUATION SIMULATION SYSTEM (AMTESS)

OBJECTIVE:

EVALUATE ALTERNATIVE AMTESS PROTOTYPES.

APPROACH:

- EMPIRICAL TEST OF TRAINING EFFECTIVENESS.
- ASSESSMENT OF USER PROBLEMS.
- ASSESSMENT OF USER ATTITUDES.

END PRODUCT:

DATA BASE AND RECOMMENDATION FOR MODIFICATIONS,
PRODUCTION AND USE.

REFERENCES

1. Waag, Wayne L., Training Effectiveness of Visual and Motion Simulation. AFHRL TR-79-72. Williams AFB, AZ; January, 1981.
2. Wheaton, George R., et al., Evaluation of the Effectiveness of Training Device: Validation of a Predictive Model. ARI Technical Report TR-76-AZ, October, 1976.

SIMTRAIN CONTRACT OVERVIEW

Stephen F. Hirshfeld
Honeywell Systems and Research Center
Minneapolis, Minnesota 55413

PRECEDING PAGE BLANK-NOT FILMED

ABSTRACT

A brief overview of the SIMTRAIN contract is presented. The purpose of the contract and the place of the fidelity workshop in this context are explained. SIMTRAIN consists of three tasks: (1) analyze and evaluate alternative analytical and empirical models for assessing training device effectiveness at key points during the device acquisition process, (2) develop guidelines for relating device fidelity characteristics to training effectiveness, and (3) evaluate training effectiveness of two alternative versions of the Army Maintenance Training and Evaluation Simulation System (AMTESS). The workshop falls primarily under task 2.

I would also like to welcome you to this meeting on behalf of the contractor team. The team consists of three companies: Honeywell, Science Applications Inc., and Klein & Associates. While originally writing the SIMTRAIN proposal, it struck us that a conference on simulation fidelity and training effectiveness would be an effective vehicle for raising and discussing issues associated with these topics. We are very pleased to have awarded the SIMTRAIN contract and to have the opportunity to assist ARI in conducting this workshop. We feel especially good about the number of participants and the diversity of services, DoD agencies, and universities represented in these meetings. We are very optimistic about the productive discussions over the next few days and their impact on the developments to be accomplished on the SIMTRAIN contract.

The contract itself is entitled "Research on Guidelines for Training Device and Simulation Development." The first task we had upon notification of award was to figure out an acronym for the program. RGTDSO didn't strike us as an optimal solution, and we were very pleased three days later at the kickoff meeting to learn that ARI had already designated it as SIMTRAIN. I would like to publicly acknowledge ARI for the creative solution. The SIMTRAIN contract started in March and is an 18-month program. We are still in the early stages of the program.

There are three distinct tasks in the contract. They are listed in Table 1. Though they are listed separately, the tasks are related, and I think insights from one will contribute to the others (Table 1). Task 1 examines both analytical and empirical devices. These methods would apply to evaluations of existing devices and also to predicting the effectiveness of alternative training device concepts. The approaches will be related to the acquisition process as characterized in the life cycle system management model. The effort will involve some development of new models and their subsequent evaluations.

The second task concerns simulation fidelity and is at the heart of the present workshop. David Baum of Honeywell will talk after I finish about the objectives of Task 2 and set the stage for the later workshop activities.

The third task is a specific application of methods/approaches for assessing training effectiveness. It concerns evaluations of prototype devices for the Army Maintenance Training Evaluation Simulation System (AMTESS). There are two competing prototypes, both offering different approaches towards the training of radar repair and automotive maintenance.

During the course of the next two days, issues will be discussed that relate to one or more of the three contract tasks. The actual structure of the workshop will be described later by Bob Hays and John Brock. I would again like to thank all of you for coming to the workshop, and I look forward to working with you.

TABLE 1
OBJECTIVES

TASK 1:

**ANALYZE AND EVALUATE ALTERNATIVE ANALYTICAL
AND EMPIRICAL MODELS FOR ASSESSING TRAINING
DEVICE EFFECTIVENESS AT KEY POINTS DURING THE
DEVICE ACQUISITION PROCESS**

TASK 2:

**DEVELOP GUIDELINES FOR RELATING DEVICE FIDEL-
ITY CHARACTERISTICS TO TRAINING EFFECTIVENESS
(WITH AN EMPHASIS ON MAINTENANCE TRAINING)**

TASK 3:

**EVALUATE TRAINING EFFECTIVENESS OF TWO ALTER-
NATIVE VERSIONS OF THE ARMY MAINTENANCE
TRAINING AND EVALUATION SIMULATION SYSTEM
(AMTESS)**

A FRAMEWORK AND TOPICS FOR EMPIRICAL
RESEARCH ON TRAINING SIMULATOR EFFECTIVENESS

David R. Baum
Honeywell Systems and Research Center
Minneapolis, Minnesota 55413

ABSTRACT

The activities and objectives of Task 2 of the SIMTRAIN program are described, along with a strategy for developing a research plan. Some problems with previous work based on hypothesis alone are presented and several empirical questions are posed. A preliminary model of factors impacting training device effectiveness is discussed, including costs. Each factor needs to be more completely conceptualized and its impact on effectiveness established through experiment.

Let me add my thanks for your participation to those who preceded me. I welcome you all to the conference. I am at the half-way point in terms of funneling our attention to what I hope will be the heart of the problem. I look forward to working with you in the next two days, to try to shape some of the approaches and research paradigms that we might pursue in the solution of the problems surrounding the determination of simulator fidelity.

INTRODUCTION AND BACKGROUND

I think we all recognize that the overwhelming issue here is how to make cost-effective decisions with regard to training device fidelity and training device features. Our bedrock question is: What is the minimum level of fidelity that is required to accomplish a particular training purpose?

This workshop is being conducted organizationally within Task 2 of the SIMTRAIN program and ARI's research program which Angelo (Mirabella) and Steve (Hirshfeld) have talked about. But, as you'll find later this morning, and the rest of the two days, the issues that we'll address go beyond fidelity, per se. Some of my remarks will try to justify that extension. I will try to give you a framework on how we are looking at the problem at this point and I'll ask you to recall that this is work in progress.

To reiterate, our goal (Table 1) ultimately is to provide guidelines for the Project Manager for Training Devices, where those guidelines will be usable by the community that has to make decisions regarding training device features and useful for making those decisions. Our approach at this point is to try to base those guidelines on valid, empirical research when possible and, where empirical research wouldn't seem suited, analytical research.

SIMTRAIN TASK 2 ACTIVITIES

These (Table 2) are the three general tasks or activities within Task 2 of the SIMTRAIN program. The first one, developing a fidelity research plan, we are still engaged in. The workshop is meant to support that development. Subsequent to this activity, we will be conducting pilot studies, and, finally, preparing a plan for the second phase of this program and developing recommendations which I would hope will spawn additional research and development on the part of the Army Research Institute.

TABLE 1. SIMTRAIN TASK 2 OBJECTIVE AND APPROACH

DETERMINATION OF SIMULATOR FIDELITY

OBJECTIVE:

**PROVIDE FIDELITY SPECIFICATION GUIDELINES THAT
ARE USABLE BY AND USEFUL FOR PM TRADE**

APPROACH:

**INSURE THAT GUIDELINES ARE BASED ON VALID
ANALYTIC AND EMPIRICAL BEHAVIORAL RESEARCH**

TABLE 2. TASK 2 ACTIVITIES

- **DEVELOP FIDELITY RESEARCH PLAN**
- **CONDUCT PILOT STUDIES**
- **DOCUMENT AND PREPARE PHASE II PLAN**

DEVELOPING A RESEARCH PLAN

The task we are engaged in now, developing a research plan, has proceeded in a number of steps (Table 3). We have reviewed the literature and we have conducted field data collection; in fact, I see some faces here of people that I have talked to out in the field. We are trying to work the problem from the top down and the bottom up at the same time. That is, we want to understand what the requirements for guidelines are in the user community and at the same time we want to understand the psychological aspects and the training impact that different levels of fidelity and different training device features have on the acquisition and transfer of task performance. Right now we are at the stage where we have identified research topic areas. The structure of our workshop is organized around those topic areas as you will hear shortly.

I should add that we will be documenting the results of this workshop in an interim report that will be published in final form sometime in October. I expect that the proceedings will be published in some detail in the form of an appendix. Each participant will receive a copy of that appendix (hopefully, that will be rather soon) and certainly a copy of the interim report when that becomes available.

THEORETICAL HIGHLIGHTS

What I wanted to do is just to review what I feel are the theoretical highlights of previous work that has been done in the area of simulation/simulator definition in terms of level of fidelity.

I think this function or set of functions here (Figure 1) will be familiar to those who have studied in this area. This is from Miller's classic work in the early 50's (Miller, 1954) and I would simply like to point out that this work was done primarily in the area of flight simulation and, as simulation is applied to a broader range of different tasks, we have to ask ourselves what the utility of this theoretical framework is for those other types of tasks.

(Question from Orlansky: Before you go too far, would you be good enough to tell us the basis on which those curves are drawn? Not the theoretical basis, the pragmatic basis, the data on which they depend.)

We have to keep in mind that this is a theoretical relationship and there are very few data that I am aware of that will allow you to identify points so that you can generate these functions.

**TABLE 3. STEPS IN THE DEVELOPMENT
OF A FIDELITY RESEARCH PLAN**

- REVIEW LITERATURE AND COLLECT FIELD DATA
- REFINE WORKING DEFINITION OF FIDELITY
- IDENTIFY RESEARCH ISSUES
- CONDUCT WORKSHOP
- SPECIFY RESEARCH PARADIGMS

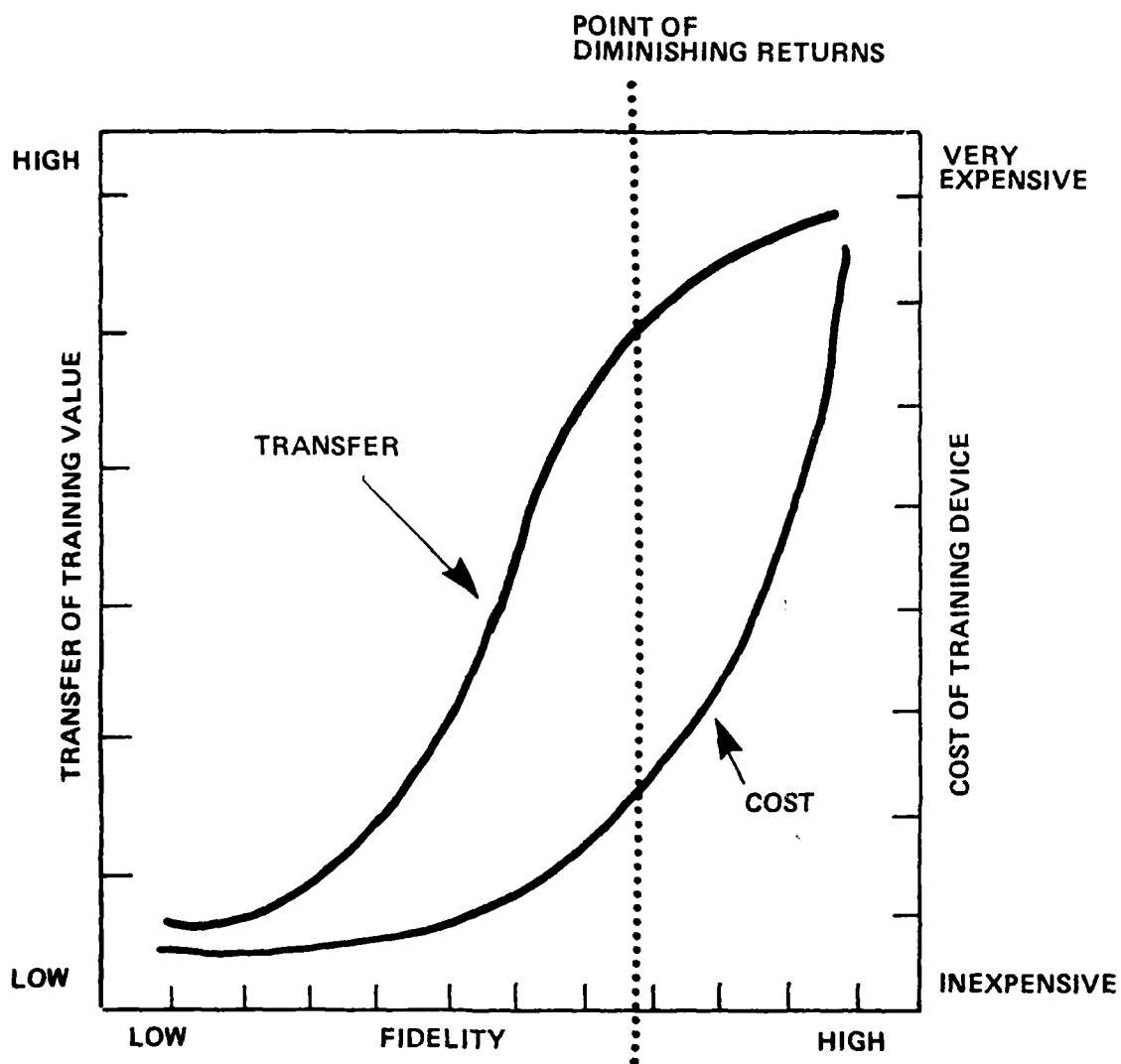


FIGURE 1
THE HYPOTHETICAL RELATIONSHIP AMONG
FIDELITY, TRANSFER AND COST
(AFTER MILLER, 1954)

I guess the transfer function that is shown here is a standard S-shape learning curve. I am not sure that we have to describe the process of acquisition and transfer.

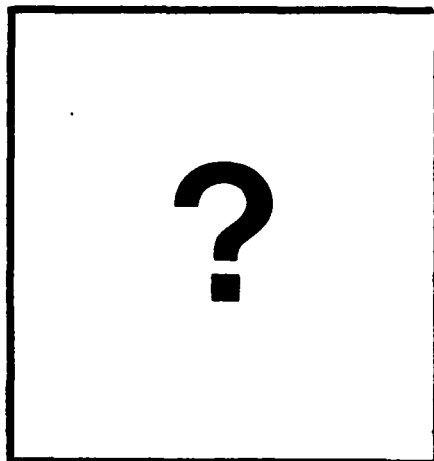
(Comment from Orlansky: That curve which was written in 1954 has been copied by everybody who has ever tried to say anything about fidelity. It happens that it was originally described as a hypothesis and Miller had no data and the repetition of that curve over and over again makes it look as if it's real. So much so that I have a little collection of about 10 people who have used it, generally using it as basis for improving its validity because they copied it from somebody else. I think it is not useful to even assume that this is an adequate theoretical explanation, quite apart from the fact that it is thoroughly unfounded and that Miller himself said it is unfounded.)

EMPIRICAL QUESTIONS

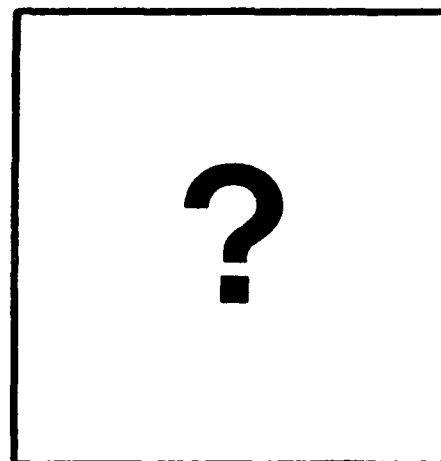
The reason I am bringing it up at this point is because I feel the exact same way; that it is simply a hypothesis. The next slide (Figure 2) shows how we might get a little bit more analytic in some respect. The previous relationship (Figure 1) dealt with fidelity as a holistic concept and, as Bob Hays will talk about later, of course the definition of fidelity has been the subject of a great amount of debate over the last 30 years. As a first cut we would like to talk about functional and physical similarity as the basis of describing fidelity and we want to raise the question: what are the shapes of those functions that will relate transfer and cost for those two different aspects of fidelity; and we would like to base the development of those relationships on as much hard data as we can so that we are dealing in a more concrete world than the hypotheses that people have propagated down through the last 20-30 years. We will initiate the collection of such data during the SIMTRAIN program.

Here is another relationship (Figure 3); this comes from Kinkade and Wheaton (1972) out of the 1972 Human Engineering Guide to Equipment Design. Again, I put this up not because it is founded on a great deal of data. It isn't; except at the low end in terms of procedures training. If there is anything that we know after the last 20 years of research, it is that you can train procedures using relatively low fidelity, and in particular low physical similarity.

But, again, this is another cut at the same problem, where stages of learning have been incorporated into the problem. To the extent that stage of learning is independent of the type of task, then I think we have to ask ourselves the question of what are the shapes of these functions in different types of tasks? Will they be the same for perceptual motor skill acquisition as they are for cognitive skill acquisition? My intuition says no, they won't be the same.



PHYSICAL SIMILARITY



FUNCTIONAL SIMILARITY

**FIGURE 2. FIDELITY MUST, MINIMALLY, BE
DESCRIBED IN TERMS OF PHYSICAL AND FUNCTIONAL SIMILARITY**

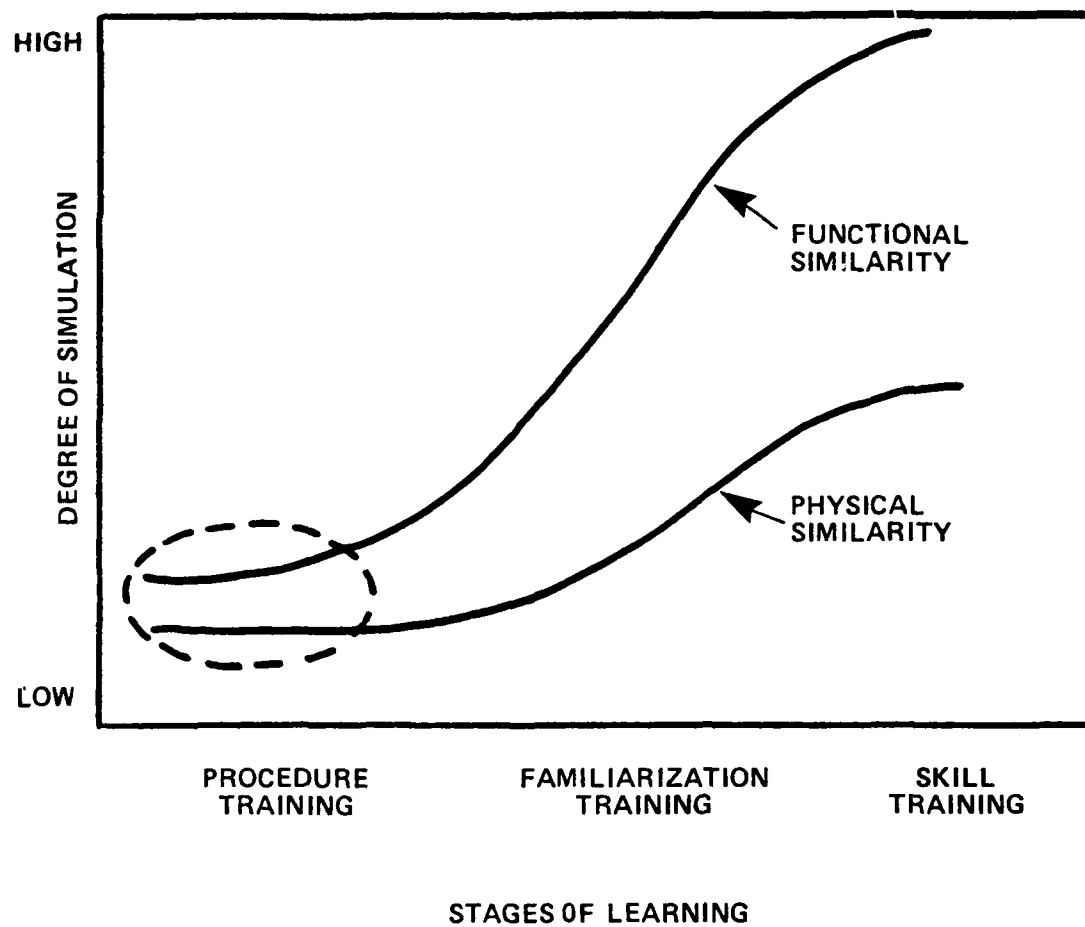


FIGURE 3
 THE HYPOTHETICAL RELATIONSHIP AMONG
 DEGREE OF SIMULATION (FIDELITY) AND STAGES OF LEARNING
 (MODIFIED FROM KINKADE AND WHEATON, 1972)

SOME KEY VARIABLES

I put this up (Figure 4) not attempting to be mathematically rigorous. This is, for me, simply a heuristic way of capturing the important components in the relationship among aspects of training devices and the types of issues we're concerned with. Cost effectiveness ratio is some figure that allows one to make comparisons among different devices or different device concepts, based on training effectiveness and cost. Training effectiveness is an umbrella term that encompasses acquisition rate and/or transfer. Transfer any way you want to measure it: in terms of first shot, savings, in terms of transfer effectiveness ratio, and so forth; where that training effectiveness is some function of fidelity, instructional features, and utilization of the device. I think we sometimes forget that last factor, i.e., utilization. These three factors clearly are not additive. Certainly they are not independent; fidelity, for example, may need to be higher than we show it psychologically necessary to be, in order to gain user acceptance; also the incorporation of instructional features may actually lower fidelity. (I guess Bob Hays will address that notion.) Also, instructional features that are provided on the devices may go unused, in which case they won't have any impact on training effectiveness.

For me this kind of captures the major topics that we need to explore empirically. I am not going to say much about cost, that is, I'm not going to talk about cost modeling, or about how one obtains cost data, but clearly it is a necessary component of this type of framework. Clearly, we need better models for each factor in the pseudo-equation pictured in Figure 4.

Because this is work in progress, I don't have any profound conclusions at this point. I certainly hope that we have some success over the next couple of days; and I look forward to having some fun. So, thank you again for coming and enjoy yourselves.

$$\text{CER} = \frac{\text{TE}}{\text{C}} \quad \begin{aligned} &= f(\text{F}, \text{IF}, \text{U}) \\ &= f(\text{Acq}, \text{Ow}) \end{aligned}$$

FIGURE 4. COMPONENTS OF TRAINING DEVICE

COST EFFECTIVENESS

REFERENCES

Kinkade, R.G., and Wheaton, G.R. Training Device Design. In H.P. Van Cott and R.G. Kindade (Eds.) Human Engineering Guide to Equipment Design. Washington, D.C.: American Institutes for Research, 1972.

Miller, R.B. Psychological Considerations in the Design of Training Equipment. WADC-TR-54-563. Wright Air Development Center, Wright-Patterson AFB, OH, December 1954.

PROBLEMS IN SIMULATION FOR TRAINING

Robert T. Hennessy
Committee on Human Factors, National Research Council

ABSTRACT

This paper outlines some general problems and issues in training that are often associated with training simulators. Some of the issues and problems discussed are: (1) substitution of simulation for operational equipment; (2) problems unique to simulation; (3) simulation as part of a training program; (4) determination of task requirements and training goals; and (5) the need for a theoretical system to relate physical (task/device) variables to performance variables by some set of intervening processes.

I am pleased that I was asked to speak about problems rather than solutions; it is certainly easier to pose problems than to solve them. I will outline some general problems and issues in training that are often associated with training simulators. This afternoon in the working groups, we'll be discussing some specific simulation issues, the "trees" in the "forest" of simulation training problems. I will discuss some general problems--the shape of the forest, if you will, rather than the trees.

The principal reasons for using simulation for training are (Table 1): it costs less than using operational equipment; hazardous tasks can be safely trained; and, it is possible to create and control conditions which are either impossible or intractable in the real world, i.e., you can achieve a given level of proficiency more quickly or attain a higher level of proficiency than is possible using real equipment.

Fidelity or realism of simulation is a problem stemming from the use of the simulator as a substitute for operational equipment. The purpose of using simulation for training, however, is not as a substitute for the real world but as a means for achieving a training goal. Simulation should be something more than a substitute for the real world.

The fundamental problem in simulation is how to achieve the most training per dollar. Achieving training effectiveness in terms of the amount of transfer of training is a limited view of this problem. The bigger problem is to get the most training per dollar comprehensively in the overall training program whether it involves simulation or not.

Most problems associated with simulator training are not any different from those in other types of training (Table 2). I think the unique problems of simulation accrue from the following: the complexity of the equipment, the number of devices involved (particularly in flight simulators), the sophistication of the equipment, (i.e., the number of potential options for simulation use in training), and the cost of simulation equipment and its use.

Simulation is only part of a training program (Table 3). Other training options are available. Classroom lecture, books, manuals, slides, movies, training devices (simulation is one of these), operational demonstration, practice on operational equipment and on the job training are all options available for achieving the training goal. I'm sure there are others; the point is that simulation is just one option within a large spectrum of possible means for achieving a training goal.

First things must be considered first--whether you are planning a training program or whether you have already decided you are going to use a simulator for some part of training. And one of the first things that must be given adequate attention is

TABLE 1

REASONS FOR USING SIMULATION FOR TRAINING

- **COSTS LESS THAN OPERATIONAL EQUIPMENT**
- **HAZARDOUS TASKS CAN BE TRAINED**
- **CREATE/CONTROL CONDITIONS**
- **MORE EFFICIENT OR EFFECTIVE TRAINING**

TABLE 2

SIMULATOR PROBLEMS ARE TRAINING PROBLEMS

- **BECAUSE SIMULATION EQUIPMENT IS USED TO FACILITATE LEARNING, MOST PROBLEMS ASSOCIATED WITH SIMULATOR TRAINING ARE NOT ANY DIFFERENT FROM THOSE OF OTHER TYPES OF TRAINING**

- **UNIQUE PROBLEMS OF SIMULATION ACCRUE FROM THE COMPLEXITY, SOPHISTICATION AND COST OF SIMULATION EQUIPMENT AND ITS USE**

TABLE 3
SIMULATION IS ONLY PART OF A TRAINING PROGRAM

- **CLASSROOM LECTURES**
- **BOOKS/SLIDES/MOVIES**
- **TRAINING DEVICES (SIMULATION)**
- **OPERATIONAL DEMONSTRATION**
- **OPERATIONAL EQUIPMENT PRACTICE**
- **ON THE JOB TRAINING**

the task requirements (Table 4). That this is a very important first step has been recognized for at least twenty-seven years. Robert Gagné emphasized this point in 1954 (incidentally, much of my presentation is based on his work (Gagné, 1954)--many issues in simulation training have not changed). The task analysis--knowing what the operational tasks are in detail--is essential. If the operational task is defined vaguely, the training goals are also likely to be vague. It is not uncommon to see training goals that are simply restatements of the general operational goals.

Training goals should be stated very specifically before attempting to outline the training program or the role of simulation in training. There should also be a good definition of the initial entry skills of the trainees and the performance criteria to be achieved through training.

Establishing specific, objective performance criteria is a difficult, but necessary, part of doing first things first. Like task requirements, performance criteria cannot be stated in general and vague terms. For example, if an operational task is to land an airplane, then stating the performance criteria as "the ability to land the airplane in a satisfactory manner" is not very helpful for designing training simulator characteristics.

Unfortunately, doing a good job of defining task requirements will not lead directly to simulator requirements. If you assume that they do, and you perform a very thorough and complete analysis, the net result is a specification for the operational equipment. The only way to take account of all possible factors that may influence training and operational performance is to include all the equipment and environmental factors occurring in the operational world.

If the simulator is viewed as a training device and not as a substitute for the real world, then it is necessary to make a big decision to answer the question: What are the critical skills that the training device is supposed to facilitate? The need for this decision is based on the assumption that certain essential skills can be more simply represented than by replicating the operational situation. That is, it should be possible to describe a skill and the conditions necessary for its acquisition in ways other than restating of the operational performance requirements.

Simulation for the training of procedural skills or continuous perceptual-motor skills should somehow capture the essence of the operational situation without duplicating it. Some loss in attainable proficiency is likely to occur if the training simulator is designed simply as a degraded representation of the operational situation. This is not necessarily bad if it is possible or cost effective to permit the desired level of proficiency to be reached during a subsequent state of training or an actual performance of the operational task.

TABLE 4
FIRST THINGS FIRST

- **TASK REQUIREMENTS**
- **INITIAL ENTRY SKILLS**
- **TRAINING GOALS**
- **CRITERIA OF PERFORMANCE**

However, the purpose of departing from duplication of the operational situation is not to achieve less costly and less expensive training devices; these are secondary considerations. the question of primary concern is: How must the characteristics of an operational task be deliberately altered to achieve the most effective training by a training device? Working on this question and the host of issues it implies is the clearest route to improving training simulation. This question will also keep psychologists and other behavioral researchers profitably employed.

What we always needed, of course, to answer the question was a theoretical system to relate physical (task/device) variables to performance variables by some set of conceptualized intervening processes. Despite the value of the development of these conceptualized intervening processes to simulation, very little research has been sponsored by the military agencies responsible for the procurement and use of simulators. Training simulators are generally procured as part of the process of acquiring new equipment. And procurement managers are more concerned about fulfilling urgent, specific needs for training simulation than supporting long term, basic research fundamental to a wide variety of simulation users. This is not a criticism; it is simply a statement of how the procurement process operates.

The remainder of my presentation will be a quick listing of problems and issues in training simulation. Few of these issues are new ones. While this list is not complete, it will give you some idea of the scope of problems that continue to deserve research. To begin, I have divided simulation problems into training issues and equipment issues. The fact that a large number of training issues exist independent of simulation equipment problems underscores the point that there are other means for getting more training per dollar than improving simulation equipment.

Problems of training strategy are concerned with the organization of the overall training program (Table 5). Problem questions include: What tasks should be trained in a simulator? Should whole task or part task training or some mix of the two be used? Should the training sequence progress in a forward or reverse chaining sequence? Should training be adaptive and progress according to the rate of skill acquisition of the students or should a fixed amount of training be given to all students? Should training be individually tailored or the same for everyone?

Other problem questions can be considered to fall under the categories of instructional methods or training procedures (Table 6): What is the role of the instructor? Should he be involved and directive or should he only provide error feedback? Should he intervene in a training event, i.e., stop the simulation, comment and then restart, or continue the simulation? Should training task difficulty be initially easy or hard? Should the difficulty level change through the course of training, and if so, should it

TABLE 5
TRAINING ISSUES

TRAINING STRATEGY: TRAINING PROGRAM ORGANIZATION

- **WHAT TASKS TO TRAIN IN A SIMULATOR?**
- **WHOLE TASK AND PART TASK TRAINING**
- **FORWARD OR REVERSE SEQUENCE TRAINING**
- **ADAPTIVE OR FIXED AMOUNT OF TRAINING TIME**
- **INDIVIDUALLY TAILORED OR SAME FOR EVERYONE**

TABLE 6
TRAINING ISSUES

INSTRUCTIONAL METHODS: TRAINING PROCEDURES

● **ROLE OF THE INSTRUCTOR**

INVOLVED AND DIRECTIVE, OR PROVIDE ERROR
FEEDBACK ONLY?

SHOULD INSTRUCTOR INTERVENE, STOP SIMU-
LATION AND COMMENT? RESTART? CONTINUE?

● **TASK DIFFICULTY**

EASY OR HARD INITIALLY?

SHOULD DIFFICULTY CHANGE? IF SO, FIXED
LEVELS OR ADAPTIVE?

● **FEEDBACK**

ERROR OR ACCURACY?

INTRINSIC OR AUGMENTED? IF AUGMENTED, HOW
OFTEN AND WHEN? ADAPTIVE OR SCHEDULED?

change among fixed levels or be continuously adaptive? Should the feedback be in terms of error or accuracy or both? Should the feedback be restricted to that intrinsic to the task performance or should augmented feedback be provided? If augmented feedback is used, how often should it occur and when? Again, should the feedback be presented in an adaptive or scheduled manner? Bicycle training wheels that are slightly above the level of the main wheels are a form of adaptive feedback. The cyclist is unaware of their presence when he is upright and doing well. He becomes aware of their presence only when he tilts to one side or the other.

Perceptual-motor skills requiring continuous performance such as in vehicle control are often taught in simulators. Psychologists often discuss continuous performance in terms of the component parts: perception (information acquisition), decision, cognitive or information handling processes, and response or motor control processes. Is it useful to consider perceptual-motor skills as a continuous cycle of behavior or are there advantages to dissecting performance into component parts for training purposes (Table 7)? In the latter case, some questions that arise include: Are the components of a skill learned at different rates? Do the components contribute differentially to skill degradation if the skill is not practiced for a period of time? If so, what are the implications for maintenance of skill training versus initial skill acquisition training? A great deal of simulation research emphasizes initial skill acquisition; but in the armed forces, maintenance of skills may be of greater concern.

There are a number of training problems which fall under the category of performance measurement (Table 8). I do not believe that the problems of performance measurement receive the attention they deserve. Without good performance measurements, you cannot do worthwhile research, make intelligent decisions about procurement of simulation training devices, decide when a trainee has achieved a criterion of performance, or determine when feedback should be given.

One issue is simply what to measure. Should just the outcome or system response be measured or should what the operator is doing, i.e., his inputs to the system, be measured? The problem of what to measure is related to the purpose of the measurement. Different kinds of measurements may be required for purposes of providing feedback, scoring performance, seeing whether a trainee meets some proficiency criterion, diagnosing training performance to discover the basis for poor performance and controlling adaptive training.

Another issue is determining or deciding on reference values for measurement. If performance error is being measured, what is the base line or zero point reference value for the measure? This is not a trivial problem and becomes increasingly difficult as the performance to be measured becomes more complex.

TABLE 7
TRAINING ISSUES

LEARNING OF PERCEPTUAL-MOTOR SKILLS

- **IS IT USEFUL TO REGARD BEHAVIOR TO BE LEARNED AS CONSISTING OF COMPONENTS?**
 - **PERCEPTION**
 - **DECISION**
 - **RESPONSE**
- **CAN THE COMPONENTS BE TRAINED SEPARATELY?**
- **ARE COMPONENTS LEARNED AT DIFFERENT RATES?**
- **WHAT IS THE NATURE OF SKILL DEGRADATION?**
- **REQUIREMENTS FOR MAINTENANCE OF SKILL VERSUS INITIAL SKILL ACQUISITION**

TABLE 8
TRAINING ISSUES

PERFORMANCE MEASUREMENT

- **WHAT TO MEASURE**
 - **OUTCOME OF SYSTEM RESPONSE**
 - **OPERATOR RESPONSES, I.E. SYSTEM INPUTS**
- **WHAT PURPOSE**
 - **FEEDBACK**
 - **SCORING**
 - **DIAGNOSIS**
 - **ADAPTATION OF TRAINING**
- **REFERENCE VALUES FOR ERROR MEASUREMENT**
- **CRITERIA FOR GOOD AND BAD PERFORMANCE, I.E. INDEX OF QUALITY**
- **WHEN IS A MEASURABLE DIFFERENCE AN IMPORTANT DIFFERENCE PRACTICALLY?**
- **IF TWO TRAINING METHODS DO NOT PRODUCE RELIABLE DIFFERENCES IN PERFORMANCE, ARE THEY EQUIVALENT? (NULL HYPOTHESIS PROBLEM)**

A third important issue in performance measurement is developing criteria for good or bad performance. Although a particular performance measure may be able to differentiate reliably between levels of performance, it is not always certain whether a higher or lower numerical value of a measure indicates better or worse performance. For example, you may be able reliably to detect differences in roll or bank angle variability among a group of student pilots but you cannot be certain that greater variability or bank angle necessarily implies worse performance. The difference in roll variability may simply be a difference in pilot control style that is of no training or operational significance. The object of performance measurement should be to provide an index of the quality of performance and not just differences in performance.

Even if it is possible to interpret a performance measurement value as indicating better or worse performance, the issue of the degree of difference, which is of practical significance, in the quality of performance remains. This is a really cost benefit question. If performance measurement is used to decide whether increased fidelity or realism of a training device is worth the cost, the remaining issue is (not simply whether increased fidelity results in reliably better performance) but whether the improvement is of sufficient magnitude to warrant the extra cost usually associated with improved fidelity.

The last issue in performance measurement that I will mention is the problem of interpreting results and making decisions when no differences in performance are found. This is a particularly difficult problem when performance measurement is being used to evaluate equipment or training methods. Suppose, for example, that two devices, indisputably different in fidelity, are used for training. When the operational performance of the students trained on these two devices is compared, no differences are found. This is a real dilemma. Do the results imply that the performance measurement methods were either inappropriate or too insensitive to discover differences in the training value of the two devices? Or do the results imply that the two devices are of equivalent value for training? Accepting the latter implication is tantamount to claiming proof of the null hypothesis. This dilemma is a problem of practicality as well as a problem in experimental philosophy.

The object of all training is to impart or improve operational skills. Transfer of training is therefore of paramount concern in simulator training research. Assessing transfer of training is difficult because it requires performance measurement in the operational context rather than in the training setting. Most of this audience is familiar with the three principal ways of measuring transfer of training: per cent transfer, the transfer effectiveness ratio, and the incremental transfer effectiveness function. An issue in simulator training is the choice of an appropriate transfer of training measure. The experimental work required to produce a per cent transfer measure or a transfer

effectiveness ratio is less than that required to produce an incremental transfer effectiveness function. Per cent transfer and the transfer effectiveness ratio basically reveal how many hours or trials of training on a device are equivalent to an hour or trial of training in the operational environment. Unfortunately, these measures do not reveal the rate of learning during the course of training. In effect, they reveal the average value per hour or trial of training. The incremental transfer effectiveness function, on the other hand, reveals the training value of each additional hour or trial of training. This last function is therefore a useful measure for deciding at what point termination of training with a particular device and continuing training with another device or in the operational setting becomes cost effective.

Because transfer of training experiments are difficult and expensive to perform, the prediction of operational performance from training performance would be very useful. However, there is always some doubt that the learning or transfer of training exhibited in the simulator during training will also be exhibited during the operational performance. I am sure that many decisions about simulator fidelity and realism requirements for training are based on the assumption that improvements in training performance lead to improvements in operational performance. A major research issue in transfer of training is to discover when this is a reasonable assumption and when it is not.

Under equipment issues, I have listed fidelity, similarity, and realism. These, of course, are the issues we have come here to discuss. It is my view, however, that these are pseudo- or superficial issues. I believe that they are identified as issues because simulation research often takes the form of comparative evaluation of different levels of simulator technology. This research attempts to relate equipment variables directly to performance with no attention to the intervening behavioral processes. The principal reason for this approach to training simulator research is a necessity for quick answers. As I mentioned earlier, training simulators are designed and procured to fulfill training support needs for some system. Research time and resources is available only for research that is perceived to be directly relevant to a particular procurement decision. Consequently, fundamental research on intervening behavioral variables relevant to simulator training has received little support. If the necessary fundamental research is performed and results in the understanding of these intervening processes, then I believe the issues of simulator fidelity, similarity and realism will disappear.

Until then the old questions, originally posed by Gagné, will continue to be current and pressing issues deserving simulator training research. The theoretical issues include: the structure of skills determinants of human variability, set, (except now I think the terms, schema and information processing strategies would replace set), motivation, and transfer of learning. Method-

ological issues include: task behavior analyses for training (I think there has been considerable useful development in this one), proficiency measurement, and criterion development. If a small portion of the resources devoted to technological improvements in simulators had been given to these theoretical and methodological issues twenty-seven years ago when Gagné first posed them, I believe we would not be discussing issues of simulator fidelity and realism today. I also believe that if the fundamental behavioral issues of training do not receive increased research attention and support, we will still be discussing issues of fidelity and realism twenty-seven years from now.

REFERENCES

Robert M. Gagné, "Training Devices and Simulators: Some Research Issues," American Psychologist, 9, 1954, 95-107.

SUMMARY OF MAJOR IDEAS FROM THORPE'S PRESENTATION
"CONSTRAINTS ON FIDELITY"

Robert T. Hays
US Army Research Institute

ABSTRACT

Important points from Major Jack Thorpe's presentation are summarized. These points stimulated discussions throughout the remainder of the workshop. Many concerned the definition and capabilities of simulation, the context of fidelity research and the constraints placed on the training community by a single "high fidelity" view of fidelity.

Major Jack Thorpe, of the Defense Advanced Research Projects Agency (DARPA), raised issues in his presentation, which provided important inputs to subsequent working group sessions. This paper presents brief summaries of the points made by Major Thorpe in that presentation.*

The first topic discussed by Major Thorpe is the distinction between engineering simulation and training simulation. This distinction is illustrated in Table 1. It is important to realize that data generated in the context of engineering simulation may not be applicable in the context of training simulation and vice versa. The two types of simulation require different research approaches, as illustrated in Table 2, and it is therefore important for researchers to maintain this distinction.

The second topic concerns the fact that training simulators exist within a total context, a total training system (Figure 1). Any change in one component will likely affect the training effectiveness of the entire system.

Thorpe's third major point is related to the above. It was to emphasize that in some cases a mix of training devices may provide the most effective training. It is therefore important for researchers to avoid looking at a simulator by itself and rather focus on the entire training system.

An important point that is often overlooked by researchers in training simulation relates to the public relations of simulator acceptance. Persons who introduce training on simulators are often viewed as "bad guys" because they are taking trainees away from what they really want to do by taking away their practice time in the real vehicle. Research is needed to determine the extent of the effect of training device acceptance on transfer of training.

Major Thorpe provided examples of visual displays of a runway during simulated landings. While most individuals would agree on which scene is of the highest fidelity, that scene did not provide the most effective training in the study described. The point is that generalizations based on impressions of fidelity or on single cases may not apply to all types of simulation or all training contexts.

The types of skills that can be trained on simulators were also discussed by Major Thorpe. There are many skills that can be trained either on the actual equipment or on a simulator. These are referred to as substitution skills because the simulator

* Original text "Constraints on Fidelity" withdrawn at request of author.

TABLE 1

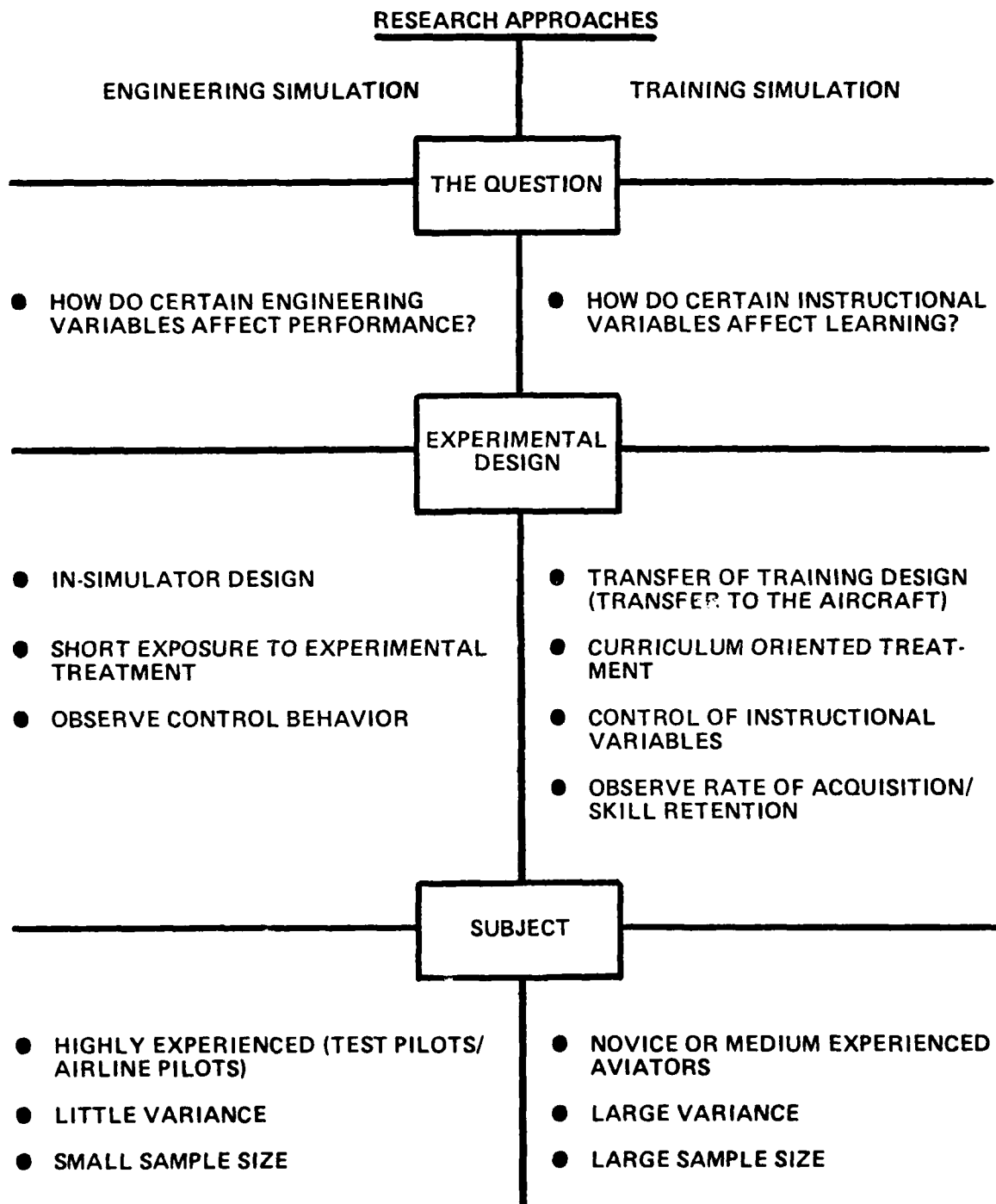
TWO GENERIC TYPES OF SIMULATION

ENGINEERING SIMULATION	TRAINING SIMULATION
<p>PURPOSE:</p> <p>To provide a testbed for studying the effects of engineering parameters on performance</p> <p>APPROACH:</p> <p>Tries to replicate all the important flight conditions</p>	<p>PURPOSE:</p> <p>To develop aircrew skills</p> <p>APPROACH:</p> <p>Identifies the skills to be trained and the types of training media which should be used (what, where, how often)</p> <p>** EXISTS ONLY WITHIN THE CONTEXT ** OF A COMPLETE TRAINING SYSTEM</p>

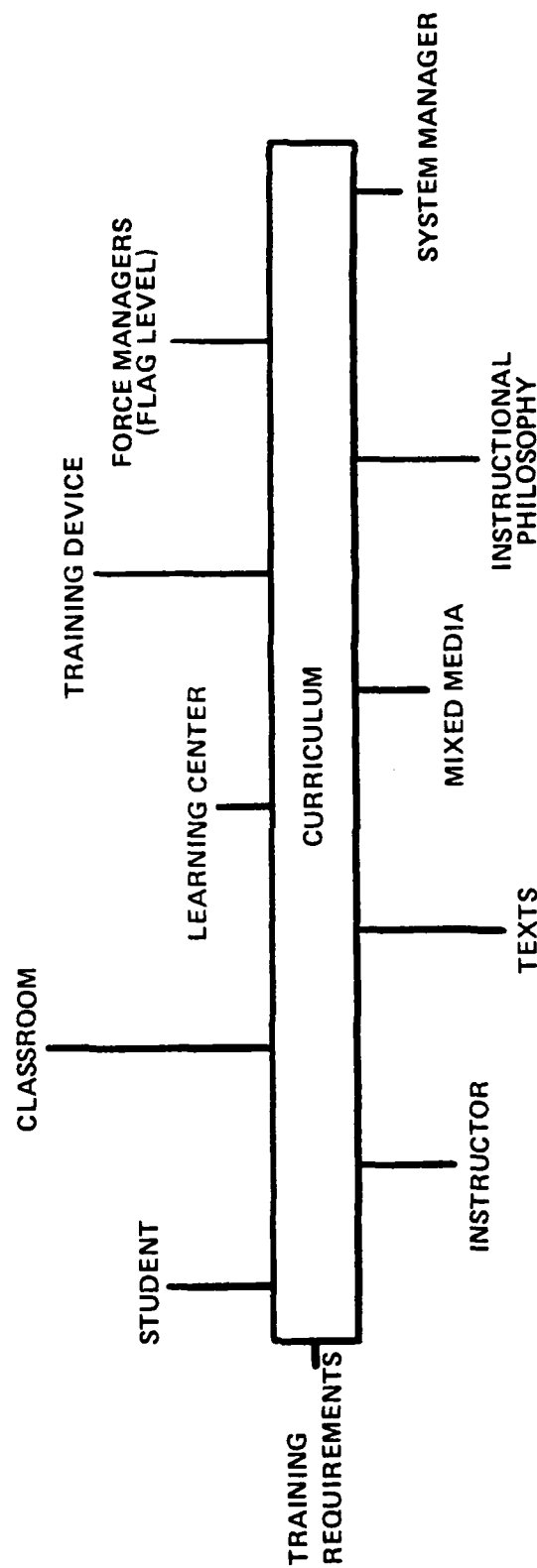
POINTS TO REMEMBER:

- THE ABOVE REPRESENT TWO QUITE DIFFERENT TYPES OF SIMULATION
- ENGINEERING SIMULATION WAS MORE ADVANCED IN 1973 PERIOD; IT HAS BEEN USED AS MODEL FOR TRAINING SIMULATION
- TRAINING SIMULATION CANNOT BE TAKEN OUT OF CONTEXT

TABLE 2



ELEMENTS OF THE TOTAL TRAINING SYSTEM



- TOTAL TRAINING SYSTEM IS EXTREMELY COMPLEX: MANY INTERACTING ELEMENTS
- ONCE ELEMENTS (E.G., THE SIMULATOR) IS TAKEN OUT OF CONTEXT, IT IS NO LONGER A TRAINING DEVICE

FIGURE 1

substitutes for the actual equipment. Some substitution skills are simple procedures while some involve complex procedures such as air to ground weapon delivery. Other types of skills that can be effectively trained via simulation are emergency procedures and tasks that are too dangerous to be trained on the actual equipment. For these types of skills, a simulator may be the only medium on which training is feasible. Finally, there are skills that are only practiced during war. These skills can be trained on a simulator although it may be difficult to evaluate the success of the training in a peace time environment.

A final point raised by Major Thorpe refers to the tendency in the training community to equate "fidelity" with "high fidelity". Thorpe feels that this attitude is counter productive because it locks us into one training strategy. It is vital that the goal of equipment design not be geared to some hypothetical level of fidelity. Rather the goal should be to design a whole training system that will make use of the most creative instructional strategies available (including low or non-fidelity). Thorpe feels that we can achieve this training strategy if we move away from the concept of fidelity as a driving force in training device design.

A FIDELITY DEFINITION AND OUTLINE
OF INTERACTIVE FACTORS

Robert T. Hays
U.S. Army Research Institute

PRECEDING PAGE BLANK-NOT FILMED

ABSTRACT

The lack of consensus in the use of the term "fidelity" is demonstrated. A definition is proposed as a "strawman" for the workshop participants. Training simulator fidelity is viewed as a description of the physical and functional characteristics of the training simulator. The interaction of these characteristics with a variety of variables must be determined if we are to understand the relationship of fidelity to transfer of training.

When I entered the field of training simulator fidelity, I was unsure of what was meant by the term fidelity. After searching the literature on simulator fidelity, I discovered that just about everybody that uses the term, uses it differently. Table 1 is not exhaustive but shows a number of different types of fidelity terms that have been used in the literature over the past few years. Not only does the literature contain a number of different fidelity terms, but we find the same terms being used to define different concepts. In Table 2, I have tried to put some of the terms together with my interpretation of what the authors meant when using the term.

The first conclusion I came to after this literature review was that the training community needed something that we could all agree on when using the term fidelity. So I began to think about just what it is that we are trying to do and the conclusion was we are trying to build training devices. We are asking fidelity questions because we want to know what the characteristics of the training devices are going to be. Fidelity can be used as a general term which describes the overall configuration of the training device. The concept of fidelity can only be useful, however, if we come to some agreement on its meaning. I propose the following definition (Table 3), for you to chew on, modify, delete, add to, or do whatever you wish during the workshop.

Basically what I assume is meant by training simulator fidelity is the degree of similarity between the training simulator and the equipment which is being simulated. It is a two dimensional measurement of this similarity in terms of (1) the physical characteristics of the simulator (what is there, what dials appear, what scenes are shown on the CRT or whatever), and, (2) the functional characteristics of the simulated equipment (what is the operator allowed to do, what information does he obtain from the displays, what can he do with the controls and to what effect). Now you might think that by limiting the definition this way, we leave out a great many concepts which may be discussed under the rubric of fidelity. I don't want to leave you with the impression that these other concepts are not useful. In fact, fidelity interacts with all of these different concepts, but it's my personal opinion that to call everything fidelity leaves us with no definition at all. To be useful in the design and evaluation of trainer configurations, the term fidelity must remain linked to the hardware. How this hardware is used, with whom, and to accomplish what purpose are questions about how the training equipment characteristics (fidelity) interacts with other variables to produce a given level of training transfer. We can't get a handle on fidelity if we apply the term to all of these interactive variables.

If you regard fidelity as the description of the characteristics of the training simulator, you've still got to see how these characteristics interact with at least the following different kinds of variables (Table 4). You've got to look at the kind of

TABLE 1
VARIETIES OF FIDELITY TERMS

AUTHOR/PUBLICATION	FIDELITY TERMS USED
KINKADE & WHEATON (1972)	EQUIPMENT FIDELITY ENVIRONMENTAL FIDELITY PSYCHOLOGICAL FIDELITY
WHEATON, MIRABELLA, & FARINA (1971)	TASK FIDELITY
MATHENY (1978)	BEHAVIORAL FIDELITY
FINK & SHRIVER (1978)	PHYSICAL FIDELITY FUNCTIONAL FIDELITY
SLENKER & CREAM (1977)	FUNCTIONAL FIDELITY
FREDA (1979)	PHYSICAL FIDELITY PSYCHOLOGICAL FIDELITY
SERVILLE (1980)	FIDELITY (TOTAL CONTEXT)
GRUMMAN (1980)	PHYSICAL FIDELITY REQUIREMENTS FUNCTIONAL FIDELITY REQUIREMENTS
HONEYWELL (1980)	SIMILARITY
TRAINING DEVICE REQUIREMENTS GUIDE (1979)	PHYSICAL SIMILARITY FUNCTIONAL SIMILARITY
INTERSERVICE PROCEDURES (1975)	FIDELITY OF JOB PERFORMANCE MEASURES
MILLER, McALEESE, & ERICKSON (1977)	DEGREE OF CORRESPONDENCE (CUES, RESPONSES, ACTIONS)
PURIFOY & BENSON (1979)	FIDELITY (TOTAL CONTEXT)
EDDOWES & WAAG (1980)	PHYSICAL SIMILARITY

TABLE 2

OVERLAPPING FIDELITY CONCEPTS

CONCEPT	FIDELITY TERMS USED	AUTHOR/PUBLICATION
PHYSICAL FIDELITY	PHYSICAL EQUIPMENT	FINK & SHRIVER (1978)
	FUNCTIONAL PHYSICAL	KINKADE & WHEATON (1972)
	PHYSICAL FIDELITY (CONTEXT)	SLENKER & CREAM (1977)
	PHYSICAL REQUIREMENTS	FREDA (1979)
	PHYSICAL SIMILARITY	HUGHES (1980)
	PHYSICAL CORRESPONDENCE	SERVILLE (1980); PURIFOY & BENSON (1979)
		GRUMMAN (1980)
		HONEYWELL (1980); TRAINING DEVICE REQUIREMENTS GUIDE (1979);
		EDDOWES & WAAG (1980)
		MILLER, McALEESE, & ERICKSON (1977)
FUNCTIONAL FIDELITY	FUNCTIONAL ENVIRONMENTAL	FINK & SHRIVER (1978)
	FUNCTIONAL PSYCHOLOGICAL	KINKADE & WHEATON (1972)
	FIDELITY (CONTEXT)	SLENKER & CREAM (1977)
	FUNCTIONAL REQUIREMENTS	FREDA (1979)
	FUNCTIONAL SIMILARITY	SEVILLE (1980); PURIFOY & BENSON (1979)
	FUNCTIONAL CORRESPONDENCE	GRUMMAN (1980)
		HONEYWELL (1980); TRAINING DEVICE REQUIREMENTS GUIDE (1979)
		MILLER, McALEESE, & ERICKSON (1977)
BEHAVIORAL	TASK FIDELITY	WHEATON, MIRABELLA, & FARINA (1971)
	BEHAVIORAL FIDELITY	MATHENY (1978)
	PSYCHOLOGICAL FIDELITY	HUGHES (1980)
	JOB PERFORMANCE MEASURE	INTERSERVICE PROCEDURES (1975)
PERCEPTUAL	PSYCHOLOGICAL	KINKADE & WHEATON (1972)

TABLE 3

TRAINING SIMULATOR FIDELITY: A PROPOSED DEFINITION

TRAINING SIMULATOR FIDELITY IS THE DEGREE OF SIMILARITY BETWEEN THE TRAINING SIMULATOR AND THE EQUIPMENT WHICH IS SIMULATED. IT IS A TWO DIMENSIONAL MEASUREMENT OF THIS SIMILARITY IN TERMS OF:

- **THE PHYSICAL CHARACTERISTICS OF THE TRAINING SIMULATOR.**
- **THE FUNCTIONAL CHARACTERISTICS (I.E., THE INFORMATIONAL OR STIMULUS AND RESPONSE OPTIONS) OF THE SIMULATED EQUIPMENT.**

TABLE 4

VARIABLES WHICH INTERACT WITH FIDELITY

- | | |
|---|---|
| ● TASK TYPE | ● STAGE OF TRAINING |
| – OPERATIONS | – INTRODUCTION |
| – MAINTENANCE | – PROCEDURAL TRAINING |
| – OTHERS | – FAMILIARIZATION TRAINING |
| ● TASK DIFFICULTY | – SKILL TRAINING |
| | – TRANSITION TRAINING |
| ● SPECIFIC SKILLS REQUIRED BY TASK | ● TRAINING CONTEXT |
| – MOTOR | – INSTITUTIONAL |
| – PERCEPTUAL | – FIELD |
| – COGNITIVE | ● INCORPORATION OF DEVICE INTO POI |
| – OTHERS | ● USER ACCEPTANCE |
| ● TRAINEE SOPHISTICATION | – INSTRUCTORS |
| – NOVICE | – STUDENTS |
| – INTERMEDIATE | ● USE OF INSTRUCTIONAL FEATURES |
| – EXPERT | |

task: Is it an operations task? Is it a maintenance task? Is it a simple procedures task? Does it involve conflict procedures? You also have to look at the difficulty of the task: Maybe you don't need a simulator to train somebody to take off a bolt. On the other hand, if you need a very precise torque on the bolt, you may need to have practice on the simulator. You may need different types, different configurations of simulators depending on specific skills that are required. You may need a different level of fidelity depending on the sophistication of the trainee or the stage of training. You also may need a different configuration of the device depending on what context you are going to use the simulator, or how it is going to be incorporated into an existing program of instruction. Are you going to write a whole new POI or are you going to somehow try to fit the simulator into what already exists? You've also got the user acceptance issue: If you have a low fidelity device, and nobody believes it's any good, and nobody practices on it, then you've got nothing. So you are going to have to determine what is the configuration of the device that will best yield user acceptance. And then you also have this last issue here, The incorporation of instructional features: Every time you put something on a simulator that doesn't occur on the regular equipment (for example, some sort of enhanced feedback or some ability to stop in the middle of a sequence and reverse or restart) you lower its fidelity because it is then different from the real equipment.

At this time, we have data on maybe 1% of these different variables. An active research program can supply the necessary data for the training community to understand the fidelity-transfer of training relationship in all its facets. To begin such a research program, it is first necessary to enlist support from many of the agencies represented at this workshop. During the workshop, we hope to address questions about how to set up this needed research in terms of the kinds of questions to ask and the types of issues to address. Please keep in mind that all of these interactive factors form the context in which our problem will be addressed throughout the workshop.

REFERENCES

- Aagard, J.A. & Braby, R.B., "Learning Guidelines and Algorithms for Types of Training Objectives", TABG Report No. 23, March, 1976.
- Adams, J.A., "On The Evaluation of Training Devices", Human Factors, 1979, 21(6), 711-720.
- Blaiwes, A.S., Puig, J.A., & Regan, J.J., "Transfer of Training and the Measurement of Training Effectiveness", Human Factors, 1973, 15(6), 523-533.
- Bunker, W.M., "Training Effectiveness Versus Simulation Realism", Eleventh NTEX/Industry Conference Proceedings, Naval Training Equipment Center: Orlando, Florida, 1978.
- Cox, J.A., Wood, R.O., Boren, L.M. & Thorne, H.W., "Functional and Appearance Fidelity of Training Devices for Fixed-Procedures Tasks", Technical Report 65-4: Alexandria, Virginia, Human Resources Research Organization, 1965.
- Crawford, M.P., "Simulation in Training and Education", Professional Paper 40-67: Alexandria, Virginia, HumRRO, 1967.
- Eddowes, E.E., "Flight Simulator Fidelity and Flying Training Effectiveness", In Proceedings of a Symposium at the Annual Meeting of the American Psychological Association, Toronto, Canada, August, 1978.
- Eggemeir, F.T., & Cream, B.W., "Some Considerations in Development of Team Training Devices". In D.E. Erwin (Ed.), Psychological Fidelity in Simulated Work Environments: Proceedings of a Symposium at the Annual Meeting of the American Psychological Association, Toronto, Canada, August, 1978.
- Fink, C.D. & Shriver, E.L.; "Simulators for Maintenance Training; Some Issues, Problems, and Areas for Future Research"; Technical Report: AFHRL-TR-78-27, July, 1978; Air Force Human Resources Laboratory, Technical Training Division; Lowry Air Force Base, Colorado.
- Fowler, R.L., Williams, W.E., Fowler, M.G., & Young, D.D., "An Investigation of the Relationship Between Operator Performance and Operator Panel Layout for Continuous Tasks", Report No. AMRL-TR-68-170, Aerospace Medical Research Laboratory, Wright-Patterson Air Force Base, Ohio, 1968.

- Freda, J.S., "Fidelity in Training Simulation", Unpublished Paper, Army Research Institute, 1979.
- Gilbert, T.F., "Human Competence: Engineering Worthy Performance", New York: McGraw-Hill, 1978.
- Grumman Aerospace Corporation, "AMTESS-1 Final Report", April, 1980.
- Holman, G.L., "Training Effectiveness of the CH-47 Flight Simulator", ARI Research Report 1209, May, 1979.
- Hughes Aircraft Company, "AMTESS-1 Final Report", March, 1980.
- "Interservice Procedures for Instructional Systems Development", TRADOC Pamphlet 350-40, 1 August, 1975.
- Kinkade, R.G. & Wheaton, G.R., "Training Device Design". In H.P. VanCott & R.G. Kinkade, (Eds.), Human Engineering Guide to Equipment Design, Washington, D.C.: American Institutes for Research, 1972.
- Lenzychi, H.P., & Finley, D.L., "How to Determine Training Device Requirements and Characteristics: A Handbook for Training Developers", Draft Technical Report TR-80, U.S. Army Research Institute, Alexandria, Virginia, May, 1980.
- Malec, V.M., "Simulation in Maintenance Training", Unpublished Paper: Navy Personnel Research and Development Center, San Diego, California, 1980.
- Martin, E.L., & Wang, W.L., "Contributions of Platform Motion to Simulator Effectiveness: Study 1-Basic Contact", Interim Report: AFHRL-TR-78-15, Air Force Human Resources Laboratory, June, 1978.
- Matheny, W.G., "The Concept of Performance Equivalence in Training Systems". In D.E. Erwin (Eds.), Psychological Fidelity in Simulated Work Environments; Proceedings of Symposium at the Annual Meeting of the American Psychological Association, Toronto, Canada, August, 1978.
- Micheli, G.S., "Analysis of the Transfer of Training, Substitution and Fidelity of Simulation of Transfer Equipment", Training Analysis and Evaluation Group, Naval Training Equipment Center, Orlando, Florida, 1972.
- Miller, G.G., "Some Considerations in the Design and Utilization of Simulators for Technical Training", Technical Report: AFHRL-TR-74-65, Air Force Human Resources Laboratory, August, 1974.

- Miller, K.E., "Simulation in Maintenance Training", Unpublished Paper: Naval Training Equipment Center, Orlando, Florida, 1980.
- Miller, L.A., McAleese, K.J., & Erickson, J.M., "Training Device Design Guide: The Use of Training Requirements in Simulation Design", Contract No. F33615-76-C-0050, Air Force Human Resources Laboratory, June, 1977.
- Mirabella, A. & Wheaton, G.R., "Effects of Task Index Variations on Transfer of Training Criteria", Technical Report: NAVTRAEQUIPCEN 72-C-0126-1, Navy Training Equipment Center, Orlando, Florida, January, 1974.
- Montemerlo, M.D., "Training Device Design: The Simulators Versus Stimulation Controversy", Technical Report: NAVTRAEQUIPCEN IH-287, Navy Training Equipment Center, Orlando, Florida, July, 1977.
- Narva, M.A., "Development of a Systematic Methodology for the Application of Judgement Data to the Assessment of Training Device Concepts", Research Memorandum 79-7, U.S. Army Research Institute, Alexandria, Virginia, May, 1979.
- Pearson, T.E., & Moore, E.O., "Analysis of Software Simulation in Computer-Based Electronic Equipment Maintenance Trainers", TAEG Report No. 57, Training Analysis and Evaluation Group, Orlando, Florida, September, 1978.
- Prophet, W.W., & Boyd, H.A., "Device Task Fidelity and Transfer of Training: Aircraft Cockpit Procedures Training", Technical Report No. 70-10, Human Resources Research Organization, Alexandria, Virginia, July, 1970.
- Provenmire, H.K. & Roscoe, S.N., "An Evaluation of Ground Based Flight Trainers in Routine Primary Flight Training", Human Factors 13 (2), 109-116, 1971.
- Purifoy, G.R., & Benson, E.W., "Maintenance Training Simulators Design and Acquisition: Summary of Current Procedures", Technical Report: AFHRL-TR-79-23, Air Force Human Resources Laboratory, November, 1979.
- Seigel, A.I., Miehle, W., & Federman, P., "Information Transfer in Display Control Systems, VI: A Manual for the Use And Applications of the Display Evaluative Index (Sixth Quarterly Report)", Applied Psychological Services, Wayne, Pennsylvania, 1962.
- Seville Research Corporation, AMTESS-1 Final Report, March, 1980.

Slenker, K., & Cream, B.W., "Part-Task Trainer for the F106 A MA-1 Operation", Technical Report: AFHRO-TR-77-52, Air Force Human Resources Laboratory, Advanced Systems Division, Wright-Patterson Air Force Base, Ohio, September, 1977.

Smode, A.F., "Human Factors Inputs to the Training Device Design Process", Technical Report: NAVTRADEVICENT 69-C-0298-1, Navy Training Equipment Center, Orlando, Florida, September, 1971.

Spangenberg, R.W., "Selection of Simulation as an Instructional Medium", Proceedings - Volume IV: Future of Simulators in Skills Training, p. 65, First International Learning Technology Conference & Exposition on Applied Learning Technology, Society for Applied Learning Technology, 21-23 July, 1976.

Swezey, R.W., & Evans, R.A., "Guidebook for Users of TRAINVICE II", Contract No. MDA-903-79-C-0428, U.S. Army Research, Alexandria, Virginia, May, 1980.

"Training Device Requirements Guide"; Army Training Support Center; PM TRADE, Orlando, Florida, & Ft. Eustis, Virginia, 5 January, 1979.

Wheaton, G.R., Mirabella, A., & Farina, A.J., "Trainee and Instructor Task Quantification: Development of Quantitative Indices and a Predictive Methodology", Technical Report: NAVTRADEVICEN 69-C-0728-1, Naval Training Device Center, Orlando, Florida, October, 1972.

Wheaton, G.R., Mirabella, A., "Effects of Task Index Variations on Training Effectiveness Criteria", Technical Report: NAVTRAEEQUIPCENT 71-C-0059-1, Navy Training Equipment Center, Orlando, Florida, October, 1972.

Wheaton, G.R., Rose, A.M., Fingerma, P.W., Korotkin, A.L., & Holding, D.H.; "Evaluation of the Effectiveness of Training Devices: Literature Review and Preliminary Model"; Research Memorandum 76-6, Contract No. DAHC 19-73-C-0049; U.S. Army Research Institute for the Behavioral and Social Sciences; April, 1976.

FIDELITY: WHAT WE KNOW AND WHO SHOULD KNOW IT

Clarence A. Semple
Canyon Research Group, Inc.

ABSTRACT

The presentation focused on user-centered issues involving training device fidelity. Users were defined as instructors, course designers, device specification writers, and students. The fundamental issue addressed was the need for improved communications between the research and development community and the user community. Issues discussed include the following: user technical education on fidelity requirements; effects of the turnover in military assignments on training device design and use; the usually meager and informal process by which instructors are trained; impacts of low physical fidelity on user acceptance and training expectations; and the need for improvements in the training device specification process.

Basically I have two objectives, and I hope to take just a few minutes. One is to make you aware of several new technical reports that are now available, some of which deal with the fidelity issue; and second to comment briefly on what I believe is a significant problem: training device acceptance. I prefer to think of the latter issue in terms of communicating some of what we know about fidelity and training to the operational user: to educate and train users in the fact that, occasionally, we do not need high realism to achieve effective training. If we don't educate users, we can expect to continue spending a lot of money for high fidelity simply to pursue physical realism.

With respect to the reports, they are a series of seven reports from a study titled "Simulator Training Requirements and Effectiveness Study" (STRES). The study was done for AFHRL/LR as part of a multiphase effort. The focus was on aircrew training and related training device issues. It was approximately a 30-month effort performed by Canyon Research Group, Seville Research Corporation and United Airlines Flight Training Center. The basic goals were to pull together the present state-of-the-knowledge in four fundamental areas: training device fidelity; instructional support features, or instructional features of such devices; principles of utilization; and life cycle costing. Because all the answers obviously are not in, an additional objective was to identify research recommendations.

The professional literature and operational experience were brought to bear in the STRES study. We cited in excess of 200 professional reports and also made approximately 30 visits to operational military users, the R&D community, airlines, and device manufacturers. So it was not just a sit at your desk analytic sort of effort, but an attempt also to incorporate "lessons learned" information. I have copies of the report citations if you are interested in them. They are available from the Defense Technical Information Center. I just want to make you aware that they exist.

The second item I want to highlight is the training device user and some of the problems I think we are facing with these people. We are facing them now, and will more and more if we depart from the operational realism concept of training device fidelity. To me, training device users are those involved in writing device specifications, developing the courses in which various training devices will be used, the instructors who implement the instruction using the device, and the students who are being trained.

The basic issue that I want to raise here is that we in the R&D community have got a very big selling, training and educating job on our hands, and one that has not really been tackled yet. I think we have to solve this communication problem with the users if the fruits of our R&D efforts are to be applied meaningfully to training device design and use.

One problem, as I see it from various programs that I've worked on, is that many device users are highly motivated but very often technically uninformed. They don't read our journals. They don't know about our technical reports or how to get them. So getting our information to them is very difficult. Also, when there are military personnel involved, they turnover every two or four years. This means a constant flow of different people coming in at different points, so there isn't a good corporate memory type of function. This also means that you have people who are forever climbing the learning curve but may never get to the top of it in a particular area. Whether it is instruction, specification writing, or course development, many users never get highly proficient before they are transferred.

User training in device utilization often is poor and unstructured. During the STRES study, we did an informal survey, asking the question: How much time in an instructor training program typically is devoted to learning to be a teacher? This is not how to fill out the forms, not the safety problems in flight or things of that sort, but how much time is devoted to learning to be a teacher? There were exceptions, but a median amount of time is about three hours, which is not a whole lot.

As we all pretty much know, learning to use training devices and simulators is a fairly unstructured process which centers around on-the-job training. "Come follow me, fellow, and I'll show you how I use this thing." Again, there are exceptions, but some examples stand out. A visit to a heavy transport training base proved quite interesting. A new instructor, an instructor under training, was being checked out in the flight simulator when the current instructor said, "Oh, by the way, I want to point out this control on the overhead console. Nobody ever told me about it. I had to find it out myself." And he turned it. It was an illumination control, and what it did was illuminate several push buttons that operated freeze, reinitialize, things of that sort. Presumably, since they land in many places around the world, not knowing about reinitialize means you would take off from Oklahoma and fly the simulator to Spain so you can practice landing there. Instructor training is very informal and I can assure you that at least the instructor training programs that I know of do not get into more sophisticated and important issues such as cognitive mediation and generalization.

I believe that training device acceptance and needs for improved user training may become even more potent issues if training devices begin looking and feeling less and less like their real work counterparts. The Army's AMTESS, for example, does not necessarily have a very high physical correspondence with its operational counterpart. AFHRL presently is studying, for example, the use of computer graphics as a simulation medium for training electronic test station operators. This experimental training device doesn't look anything like the actual test

station. So we are trending into some areas where we depart from what the user typically thinks of as being high fidelity. I believe we are headed for trouble unless we approach it properly.

One thing that I have found that instructors strongly fear is that a departure from high physical fidelity and realism will lead to negative transfer of training. Not that no training will occur, although there is some concern with that, but mostly that negative training will occur. They are quite concerned that enough physical differences between the device and the real world situation almost universally means practicing on the device is bad. I think many of those fears are unfounded, but this also remains an issue that has to be dealt with.

With respect to specification writing (I make these statements rather generally and they certainly don't apply universally), we have seen one approach used over and over again. Assume you are a military person who has to make inputs to a training device specification. What do you do first? Typically, you go find an old specification for a somewhat similar device and use that as your model. In other words, you steal from history. If you stop and think about this, it is quite contrary to R&D, and it is quite contrary to state-of-the-art, which, to me at least, implies advanced technology and thinking.

As Jack Thorpe and Bob Hennessy mentioned before, all too often in the device specification process, too little attention is given to the particular and specific training tasks and training objectives that the device is to be used to train. Very seldom is any consideration given to the design of the courses in which the training device is to be used when the device specification is being developed. I may be an old fuddy duddy, but I always have found it confusing to specify training device characteristics when I don't know the particulars of what the device is to be used for.

In summary, I believe we have a big user-oriented selling and communication problem, along with associated user training needs. I see the problem as becoming particularly nasty as some of the fidelity research shows that we can depart from the realism concept of such devices, at least in some cases. And I think it is up to us in the R&D community, as we discuss fidelity issues, to also begin thinking very seriously about how we are going to educate/train the operational user to accept some of the work that we have done which impacts on their concepts of what makes a good training device.

REFERENCES

- Caro, P.W., Shelnut, J.B., and Spears, W.D. Aircrew Training Devices: Utilization. AFHRL-TR-80-35. Wright-Patterson AFB, OH: Logistics and Technical Training Division, Air Force Human Resources Laboratory, January 1981.
- Allbee, K.E., and Semple, C.A. Aircrew Training Devices: Life Cycle Cost and Worth of Ownership. AFHRL-TR-80-34. Wright-Patterson AFB, OH: Logistics and Technical Training Division, Air Force Human Resources Laboratory, January 1981.
- Prophet, W.W., Shelnut, J.B., and Spears, W.D. Simulator Training Requirements and Effectiveness Study (STRES): Future Research Plans. AFHRL-TR-80-37. Wright-Patterson AFB, OH: Logistics and Technical Training Division, Air Force Human Resources Laboratory, January 1981.
- Spears, W.D., Sheppard, H.J., Roush, M.D., and Richetti, C.L. Simulator Training Requirements and Effectiveness Study (STRES): Abstract Bibliography. AFHRL-TR-80-38. Wright-Patterson AFB, OH: Logistics and Technical Training Division, Air Force Human Resources Laboratory, January 1981.
- Semple, C.A., Simulator Training Requirements and Effectiveness Study (STRES): Executive Summary. AFHRL-TR-80-63. Wright-Patterson AFB, OH: Logistics and Technical Training Division, Air Force Human Resources Laboratory, January 1981.
- Semple, C.A., Hennessy, R.T., Sanders, M.S., Cross, B.K., Beith, B.H., And McCauley, M.E. Aircrew Training Devices: Fidelity Features. AFHRL-TR-80-36. Wright-Patterson AFB, OH: Logistics And Technical Training Division, Air Force Human Resources Laboratory, January 1981.
- Semple, C.A., Cotton, J.C., and Sullivan, D.J. Aicrew Training Devices: Instructional Support Features. AFHRL-TR-80-58. Wright-Patterson AFB, OH: Logistics and Technical Training Division, Air Force Human Resources Laboratory, January 1981.

EFFECTIVENESS ISSUES

Rohn J. Hritz
Science Applications, Inc.

ABSTRACT

Several topics concerning the determination of the effectiveness of simulator configurations are discussed. The point is made that fidelity may not be an overall thing. It may be necessary to make fidelity decisions on each component in a device. Another point concerns the fact that decisions are now being made about device design. We need to determine how this process works and how it can be improved. We also need to improve the communication between different individuals at various stages of the ISD process. Finally, we need to develop methods of evaluating devices when there are no competing devices to compare. References are provided for current efforts in this area.

I really don't have much to add to what has already been said, but maybe I can bring out a few points. If we take a look on how trainers are designed right now, we typically do a front end analysis and the front end analysis is geared at putting a whole training system together and not only designing a trainer which may be part of that training system. We go through and decide what task needs to be trained and make a decision on what tasks need to be acquired on some sort of simulator or trainer. If a trainer is needed, then we sit down and make some decision about the requirements of that trainer, particularly its level of fidelity, which is not an overall thing, each component on the trainer has to have a fidelity decision made for it. When we decide what instructional features are supposed to be on that device, etc., one of the things I think that we forget is that we're making these decisions now, today. We are designing trainers, trainers are being used so we are making these decisions. Somebody out there must have decision making models for making these things. I think often what we fail to do is to trace back to those decision making models, once we do an evaluation study. Typically, what occurs is that a Tech Sergeant or someone at that particular level is usually responsible for doing a front end analysis. He makes the decision to have the trainer. He makes the decision on what characteristics the trainer is supposed to have. Somehow those requirements, those training requirements, get translated into a procurement specification. Now, there is a lot of loss. When that thing goes from the training requirements to the procurement specification, there are some things lost. As the trainer is being built there is some acceptance testing going on or some kind of formative testing and when the trainer is built we also do some evaluation studies. What we fail to do is go back and take a look at those studies and trace them back to the decision models that were used in making fidelity and deciding what instructional features the device has. I think that is an important issue.

I think the other important issue is that a lot of times comparative studies are done, and the comparative studies are done between training systems. Is this training system better than this training system? That's kind of nice but we are at the point now where we don't have alternate training systems. It is getting too expensive to build two alternate training systems for the same kind of training. We don't do that anymore. So I think we have to come up with ways, in the practical world, where we can begin to do effectiveness studies when there is no competing or alternate training system. That is one thing I think this group ought to work with today and tomorrow.

I guess the other objective I have, and I will keep my comments short, is that we have been working with HRL at Lowry AFB and we have put together an ISD procedural handbook and I just want to make you aware that these documents exist. The procedural handbook addresses some of these issues. What tasks need to be trained, what tasks need to be trained on a trainer on a simulator, as well as deciding what requirements that particular

trainer ought to have. We have also put together a system where you can trace back the decisions made by that ISD handbook after the trainer has been built. What occurs is that the ISD procedural handbook is used by an ISD analyst, he records those training requirements in what we call an ISD derived specification. That specification goes to what is called the System Program Office, the SIMSPO, those training requirements are supplemented with engineering requirements and go into a procurement spec. As the device is being built and as acceptance testing is going on, it is our hope that you will be able to trace decisions, particularly fidelity level decisions or decisions to have instructional features, that are either good or bad, back to the model or decision making procedures that generated those requirements. We have not yet field tested that particular approach. But those documents are available if you are interested.

REFERENCES

- Hritz, R.J. Behavior Taxonomies and Training Equipment Design: A Literature Review and General Model. Valencia, PA: Applied Science Associates, Inc., April 1979.
- Hritz, R.J., Harris, H.J., Smith, J.A., & Purifoy, G.R., Jr. Maintenance Training Simulator Design and Acquisition: Handbook of ISD Procedures for Design and Documentation. Valencia, PA: Applied Science Associates, Inc., December 1979.
- Hritz, R.J., Purifoy, G.R., Jr. Maintenance Training Simulator Design and Acquisition: ISD-Derived Training Equipment Design. Valencia, PA: Applied Science Associates, Inc., December 1979.
- Hritz, R.J., Purifoy, G.R., Jr. Maintenance Training Simulator Design and Acquisition: ISD-Team Training Course Outline. Valencia, PA: Applied Science Associates, Inc., December 1979.
- Hritz, R.J., Purifoy, G.R., Jr. Maintenance Training Simulator Design and Acquisition: Model Training Equipment Functional Specification. Valencia, PA: Applied Science Associates, Inc., January 1980.
- Hritz, R.J., Purifoy, G.R., Jr. Maintenance Training Simulator Design and Acquisition: SPO Orientation Training Outline. Valencia, PA: Applied Science Associates, Inc., February 1980.
- Hritz, R.J., Purifoy, G.R., Jr., Harris, H.J., & Smith, J.A. Performance Technology in the Armed Forces: New Techniques for Maintenance Training Simulator Design. Paper presented at 18th Annual Conference of National Society for Performance and Instruction, April 1980.
- Hritz, R.J., Purifoy, G.R., Jr., & Smith, J.A. Maintenance Training Simulator Design and Acquisition: Prime Development Specification for Maintenance Training Simulators. Valencia, PA: Applied Science Associates, Inc., April 1980.
- Purifoy, G.R., Jr. & Benson, E.W. Maintenance Training Simulators Design and Acquisition: Summary of Current Procedures. Technical Report: AFHRL-TR-79-23. Technical Training Division, Lowry AFB, Colorado, 1979.

DEFINITIONS IN SIMULATION

Worth Scanland
Office of the Chief, Naval Education and Training
Pensacola, Florida

PRECEDING PAGE BLANK-NOT FILMED

ABSTRACT

Several points are made about the area of training simulation research. The matter of definitions is critical. It's very important that there be designed a glossary to which we can all pay homage and agree. More attention should be paid to the area of training through stimulation, simulation with a "t". This is particularly true when actual equipment must be used but it is not feasible to operate it in a real world environment.

I really don't have a great deal to say. My suggestions to you, John (Brock), I think it was you, when you asked if I had a particular topic which I thought ought to be included to these proceedings was that the matter of definitions is critical. Some of us here were at a SALT meeting about a week ago on front end analysis as it applies to simulators and the issue of definitions arose because it was so apparent, certainly to me, and I hope to others, that we are inclined to speak the same words with a very different meaning. And in this kind of esoteric domain in which we are now dealing, I believe it's very important that there be designed a glossary to which we can all pay homage and agree. I looked up some of the terms just for fun in my office. I went to Webster which is ordinarily considered to be a reasonable authority on definitions. He says that to simulate means to pretend, to feign, to counterfeit, to give false indication. Now, very little of that is meaningful to us. I looked in a number of Navy publications where each one differs in its definition relating to training devices, technical training equipments, training aids, and simulators. One of my colleagues who has recently done a study on simulators in maintenance training, named Jesse Orlansky, has a nice, simple definition of a simulator in his glossary. It says a simulator is a device on which simulation is implemented. I think that is a very nice definition. I want to congratulate him on it. And it is perhaps indicative of the need for simplicity in our definitions. I liked Bob Hays' definition of fidelity. I had written one, which I am amazed to see, is almost verbatim a match for his, in which I said it is the extent to which a training device, including simulators, replicates the physical characteristics and environment of the operational system for which the training is being provided. Now three or four speakers before me have spoken to the subject of fidelity and I guess my overall impression is that there must be 50 different meanings to the term depending somewhat on the adjective you put before it. I need, and I hope you do too, a better definition of evaluation as it relates to training devices and simulators. At a meeting recently, I proposed that we use that term to measure the degree to which the device meets the engineering specifications and that the term validation might be useful then to describe the degree to which the device leads the student to the achievement of his learning objectives. Jack Thorpe raised an interesting point, from my perspective, in his declaration that good training requires a sort of melding of experiences in real life and other kinds of instruction including that on simulators. I hope that's a fair summary of what you said Jack, in that stairstep thing you had, which led me to think that we haven't talked at all, at least not yet, in this meeting about stimulation, simulation with a "t", that I believe has a great deal of prospective merit down the road, particularly wherever the simulation is too expensive and you must use the operational equipment but you cannot afford to really operate the equipment in its real world environment. We have ships in which there are as many as 34 consoles in one control room and that requires a great deal of individual as well as team training. It is almost as expensive to do it on the beach as it is aboard ship, but it can't

be done aboard ship, of course, unless a scenario is played and signals and inputs are made to the sensing devices that are very important and essential part of the total weapon system. And therefore, it seems to me there is a great deal of room for research in how we can introduce software into these complex devices, which all now depend upon computers, in such a fashion that we could put the operator on his real equipment but we don't burn up any kind of fuel running the ship or aircraft around and so forth.

IS SIMULATION FIDELITY THE QUESTION?

William E. Montague
Navy Personnel Research and Development Center

ABSTRACT

A review was done of recent work on the analysis of learner's representations of complex physical systems. This was used to introduce the concept of mental models to show that on one hand they are an important part of what must be considered in designing simulation, and on the other hand that faulty mental models of what simulators should be like ignore almost everything we know about how to teach people to do complicated tasks. The paper suggests that simulators for training must be designed using specifications that include characteristics that promote learning and understanding. Sometimes it may be necessary to violate physical and temporal fidelity to promote learning. A development effort is recommended to operationalize guidelines for specifying and designing instructional task fidelity. An effort is needed to promote and monitor their use.

INTRODUCTION

Mental Models

When the word "fidelity" is applied to simulators and training devices it means different things to different people. At a general level it is assumed that high fidelity is necessary for a device to be an effective trainer. "High fidelity" is often defined in terms of whether the device generally looks, feels, smells and runs like the actual object being simulated. The problem is that this approach is too general. Training that the device is to support, is left only vaguely specified. The design specifications do not include the actual training requirements to guide performance measurement that documents learning, nor do they specify the conditions of use necessary to bring about training. This seems to be due to a limited "mental model" or conception of how to specify the design requirements.

The notion of mental models is a useful way to discuss some issues relevant to the question of the fidelity of simulations for effective training and maintenance of competency. The mental models notion is really not new, but is receiving considerable recent attention among cognitive scientists (e.g., Johnson-Laird, 1980, de Kleer & Brown, 1980). The notion has considerable heuristic value that can be used in the current discussion.

Basically, the idea is simple, a person makes use of an internal model of the world to reason and explain things about the world.

If the organism carries a "small-scale model" of external reality and of its possible actions within its head, it is able to try out various alternatives, conclude which is the best of them, react to future situations before they arise, utilize the knowledge of past events in dealing with the present and the future, and in every way to react in a much fuller, safer and more competent manner to the emergencies which face it. (K. Craik, 1943, cited by Johnson-Laird, 1980).

Mental models are composed of autonomous objects associated topologically with others, rules for their interaction that allow them to be "run" in one's imagination or mind's eye and the outcomes assessed. They assist human reasoning by producing explanations or justifications of complex system behavior. They are mnemonic devices or learning devices (Hollan & Williams, 1981). Thus, they aid in the apprehension of how systems work and provide a strong means for generating expectancies about how things are done, should be done, or the consequences of certain actions. It is important to realize that mental models are powerful analogical devices humans use in learning about their world.

There are two main points to emphasize about this notion of mental models. First, people develop mental models through the interaction with physical and social systems. Thus, we must consider how to present material and simulation to best develop effective training. The principles for designing and critiquing instructional presentations (simulations) need to be developed (de Kleer & Brown, 1980). Second, since adults may have strong models replete with errors or "bugs", they can also inhibit new learning, or changes in systems. So on the one hand mental models are used by individuals to reason about systems qualitatively and "handle" their world, and on the other they may provide a barrier to change or correction.

In training with simulation, we are attempting to help people build up their representation of the physical world to be able to operate equipment effectively, and we must be able to design training and training devices to allow the most effective buildup of appropriate mental models. Many of the tasks to be trained are procedural and involve decision about reporting and classifying signals, or anticipating outcomes resulting from changes in a system. Simulation for these tasks must be done in a way that maps into what the trainees already understand as well as mapping the critical aspects of the terminal task to be learned.

In addition to needing to consider the mental representation of the learner to design simulators, we need to be concerned with the mental models of training simulators possessed by the managers of the personnel and training system. From the point of view of the education/training psychologist, these people are laymen. They don't take into account the need for performance measurement, practice, and feedback needed for learning. Thus, it has been difficult to introduce systematic changes into how they specify requirements for training devices.

PROBLEM

People's Limited Mental Model of Simulation

When one discusses "fidelity" of simulation what most people think of is how much physical resemblance or isomorphism there is between the simulated and "real" task situations. This seems to be the reason that we work so hard to create simulators that look, feel, smell, bounce around and otherwise seem to operate like the real equipment. A simulated plane's cockpit must be indistinguishable from the real thing. A simulated power plant must generate steam power. That seems to be the philosophy guiding the construction of simulators. We know that this is costly, but assume that it is worth it because it is so real. Fidelity is a close synonym of "realism" in this usage. Realism may be useful for some things at some times, but it may not be providing the best training situation in many, even most instances. It may slow learning down considerably.

This idea probably describes most laypersons' model of what an effective simulator or training device should be. Mental models like this influence expectancies, and therefore, what they buy or build when they are assigned the task of developing training devices. Since most training system managers are "laymen" when it comes to knowing how to train, the characteristics of simulation needed for effective and efficient training are often not specified and required. Similar perspectives about classroom instruction also allow ineffective methods to persist and slow down the rate of acceptance of more effective teaching techniques (Montague, Ellis & Wulfek, 1981). Everyone thinks they know how to teach because they went to school. Similarly, it is easy to assume that a realistic training device will be an effective trainer. Such misapprehensions are difficult to overcome. In these cases the problem is that the training/teaching is approached in too general a fashion. No specification of training requirements and objectives is made with the consequential failure to define performance criteria needed to judge the success or failure of trainees.

The State-of-the-Art

Let's take a moment to look at the consequences of this limited mental model in action. Let me describe what seems to be the state-of-the-art of designing simulators. "State-of-the-art" in this case is defined in terms of the simulators being acquired to assist in training sonar operators to go into use sometime in 1985. I will outline the surface sonar operator's task, describe the training briefly, and describe the concepts behind the simulation to be used for training. From this description we will be able to evaluate the quality of these simulators for training on the basis of what we know should be included to make training effective, and to understand the effect of a more limited, erroneous but prevalent mental model on training device design.

The tasks of detecting and then classifying signals on various displays are the primary, and mission-critical tasks required of sonar operators. They are complicated and difficult. Interestingly, they are not emphasized in training. Operator training concentrates on the somewhat simpler tasks of searching, tracking and localizing possible targets (Mackie & Schult, 1981). The trainers used for teaching operators historically use simplified detection and classification tasks, sometimes eliminating them completely as tasks performed in training. There are a few, rare, operator trainers developed recently that present recorded or good simulated signals. However, they provide no assessment of errors and appropriate feedback. Thus, the use of these devices in schoolhouse training is unlikely to provide effective training for operators that would allow them to perform well once they are assigned the tasks of detecting and classifying signals on the job.

The job itself does not provide systematic experiences that will develop the detection and classification skills. At sea, most ships are in areas that have few contacts of military significance. In addition, a sonarman is assigned many other tasks to do beside sonar watchstanding. Even when a ship is regularly included in "Exercises" there is no assurance that each operator will receive systematic exposure to the military targets so that they will develop the high skill levels needed to perform the complex pattern recognition tasks involved. Novice operators are not assigned systematically during exercises, and there is no sound basis for rapid, thorough feedback to the operator regarding either successful or erroneous performance. So we have neither formal schooling nor the job experience providing conditions for the systematic development of these skills so necessary to readiness and the Navy mission (cf., Mackie & Shultz, 1981).

Naturally, these problems have not gone completely unrecognized. However, it is important to examine the quality of the solutions proposed. By 1985 the navy is supposed to deploy an integrated passive sonar system to many of its ships. This system integrates three sensor systems. The operational complexity of this system and the tasks of operating it have been recognized, and the Navy is procuring a few, very expensive dynamic team trainers to provide teams with "practice." The design concept for this integrated system is captured in the words: "high fidelity." The actual operational equipment is simulated, giving the equipment the ability to show targets as they actually might occur, permitting their detection and classification, and allowing realistic execution of other phases of ASW operations. On the surface this sounds good, but the cost restricts the number of units to be purchases. Since such simulation works in "real" time, few people can be trained on the system restricting its use to more "advanced" training in preference to initial operator training or refresher training. In addition, such systems require many, highly trained personnel as instructors who provide feedback to team members. How to manage this is left to those individuals.

In order to take care of more fundamental operator training, the Navy has a series of Basic Sonar Operator Trainers (BSOT) under development. These are designed around the mainframe cabinets and represent operator consoles of specific sonar systems. Notice that "high fidelity" in this instance refers to the physical consoles themselves along with their knobs and buttons. The trainee can manipulate console controls in response to static information presented by means of carousel slide projectors holding 80-120 slides. They can learn how to set up equipment, and learn certain procedures on such devices. Training in the critical skills of detecting and classifying signals is rudimentary.

If this example is representative of the state-of-the-art in simulators for training, we can detect two design approaches. First, operational training devices are designed that provide high

fidelity presentation of the actual equipments to be operated and the signals to be handled. These devices are expensive, work in real time, and provide exercise for a team of operators on the various subsystems. This is the case in the example described as well as in propulsion engineering, aviation, and other complex weapon systems. In these systems there is no provision for measuring the performance of various operators so that the adequacy of their performance can be assessed, and their errors diagnosed. Nor is there provision for systematic exposure to critical or difficult signals and tasks appropriate to operator deficiencies or lack of skill. These systems make no provision for rapid and comprehensive feedback.

To take care of basic operator training, the emphasis is on a trainer that teaches people how to set up and calibrate the equipment and carry out some of the simpler procedures. But the interactive capability is poor. The important operator tasks of detecting and classifying receive only cursory attention. The presentations are static, their number severely limited by the use of slides. The dynamic character of the actual task to be performed on the job is not well represented. In sonar, much of the critical target information is time dependent. Important aspects of performance such as detection latency, reporting accuracy, and classification latency and errors are not taken. The "state-of-the-art" in trainer design seems to ignore almost everything we know about how to teach people to do complicated tasks.

What are the characteristics that should be present in devices that are to be used for training, and how might they be used prescriptively in designing simulation?

Simulators That Teach

Expensive, physically isomorphic simulators may not be good instructional devices, or exercisers for a number of reasons: they are large, events happen over long time periods (as they do in the "real" world) and provide relatively little practice. They are so big or complex that the student cannot "see" the result of some action on the operation of the system. The system may not provide good feedback about errors. In other words, they violate what we know about how to teach people.

What is required for training simulators/devices is that they must provide the cues, the opportunities for trainees to respond, make and correct errors, and observe the consequences of their actions. Simulation that teaches well may violate the physical and temporal isomorphism, and its adequacy in training may be due to that fact. This idea is counterintuitive to the naive view that the most important design goal is to provide high physical and stimulus fidelity. Simulators that are effective trainers need to simulate faithfully the tasks to be learned in a way that teaches students to think about the problems they confront in

using the equipment and build appropriate mental representations of the system and the way it operates. Thus, they learn to operate it well, can respond to emergencies and anomalies in the system effectively. In order to do this, a simulator needs to show, perhaps quickly, perhaps not, what changes in a system occur because of certain actions. It should allow frequent and rapid practice of procedures to be learned, allow and provide corrective feedback for errors so that the errors can be eliminated. The real equipment, as well as physically and temporally close simulations of the real equipment may not, and often do not provide this instructional environment.

It may also be true that simulators intended for training novices and those intended to provide extensive practice or retraining for moderately competent people may need to be designed quite differently. The need for extensive corrective feedback is substantially different for these groups. Novices need extensive guidance, and precise corrective explanation of their errors and the reasons for them, while already trained individuals may need refreshment, or may need to broaden their knowledge base for situations or signals to be encountered.

The primary idea is that simulation needs instructional "task fidelity" (Semple, et al, 1981) and that requires determining training device cue and response capabilities needed to support learning and practicing tasks that lead to or support the operational performance required. A major problem is that such determination is based on learning task analysis procedures which are not well developed, especially for complicated tasks involving decision making and problem analysis. Better techniques are needed so that training devices can be designed to support trainee learning and practice, rather than simply mimicking operational system characteristics and uses. To do this, the final skills to be learned need to be specified in terms of task performance, the performance conditions, and performance standards. Also, the steps involved in apprehending the skill need to be identified.

The following section of the paper attempts to outline some tentative guidelines for designing training devices in terms of training task fidelity. "Fidelity" is directly linked to the specification of the knowledge and experience required to support specific performance. The focus is on identifying specific training objectives, and on arranging the process of training to accomplish them.

In Table 1 are listed some major general goals of training devices. Any simulation is trying to accomplish these things. They try to build up experience that will aid the student in the performance of the real task. For the most part the goals in Table 1 are self explanatory. The first item, however, needs a little more explanation. A major goal is to help people build a mental model of the system and its functioning and of the consequences of their actions in the system. In complex systems

TABLE 1

MAJOR GOALS OF TRAINING DEVICES/SIMULATORS

- **BUILD UP IN THE LEARNER'S MIND/MEMORY A REPRESENTATION OF THE TASKS TO BE PERFORMED AND HOW THE SYSTEM FUNCTIONS. THE GOAL IS TO BUILD UP THE TRAINEE'S UNDERSTANDING OF THE SYSTEM AND HIS ROLE IN IT SO THAT HE CAN MAKE APPROPRIATE INFERENCES ABOUT WHAT TO DO UNDER VARIOUS CONDITIONS, AND WHAT WILL HAPPEN IF HE DOES SO.**
- **PRODUCE SMOOTH, ERROR FREE, EXECUTION OF NEEDED PROCEDURES/RESPONSES**
- **TEACH IN A SAFE, LESS COMPLICATED ENVIRONMENT, AND PERHAPS SHOW THE CONSEQUENCES OF UNSAFE ACTIONS.**
- **PROVIDE FOR THE DEVELOPMENT OF SKILLS.**
- **PROVIDE TUTORING (ESPECIALLY INTELLIGENT) THAT GIVES HINTS, COACHING, WARNINGS, EXPLANATIONS**

we would like the learner to develop an accurate understanding of the system that will allow him or her to infer what will happen if some action is taken, or what produced some problem or symptom. In part, limits in our ability to specify and train exhaustively on complex systems require this understanding.

Rules or Guidelines for Designing Simulations for Training

What makes a good simulation? There are two main characteristics: they must be recognizable representations of some aspect of objects or systems, and the representations help us to understand, by manipulating, enjoying, or predicting the behavior of these things. How can we deliberately contrive simulations to facilitate learning, to provide insight into complex system functioning?

To be powerful for training, simulators must not only simulate reality, and be useful in anticipating and controlling that reality, but must effect or change, or have impact on those manipulating them. Effective simulators involve the learner; he "flies" a link trainer and sweats, or gets upset when he views injustice in a film, or he detects his error in opening a valve too soon that causes a casualty and is chagrined, but recognizes that he could have caused a real disaster.

But, as we indicated above, physical similarity or identity may not allow the most effective training (i.e., the most involvement). For example, Johnson (1981) recently demonstrated that limited fidelity simulation (a paper mock-up) was more effective than the real equipment in producing performance resistant to forgetting (over 60 days or so). This effect was produced because the limited fidelity simulation required trainees to rely more on their memory during learning. Similar findings have been reported by a number of other researchers (see Johnson, 1981 for other citations).

The essence of simulation is placing the learner in a sufficiently realistic situation that requires the same sequence of inquiries, decisions and actions appropriate in the actual situation. Each of these activities triggers appropriate feedback which may or may not alter the situation, but which can be used for subsequent decisions about further actions that may change the nature of the problem. Therefore, a prime characteristic of effective simulation is in terms of how it mimics the critical characteristics of the system, its problems and its functions, from the point of view of the learner.

The analogical relationship between the simulation and the real task function is primary. People build up representations by observing the effects of procedures they apply (at least active practice is better, although passive practice can be effective in some cases). So design must make this observation possible. Often, realistic simulators don't make this observation easy, and

some even impossible (e.g., in a steam plant simulator, the effects of actions cannot be seen for long time periods, or without physically moving somewhere to observe events or situation readouts on gauges).

Teaching by simulation is teaching by analogy. In such teaching, the mapping between a learner's existing knowledge structures and the training (simulation) determines learning difficulty and how and what is learned (J.R. Anderson, et al, 1981; Bott, 1979; Rumelhart and Norman, 1981). Similarly, the mapping between the training/simulation and the actual task should determine learning and the adequacy of the resulting performance. Therefore, "analogical correspondence" (Riley, 1981) between the actual task situation and the representation is a primary dimension in designing simulation. The problem, of course, is defining that dimension. Gentner (1980) proposed a structural characterization of good scientific analogies where the mapping between the analogy and the scientific explanation/model/task can be specified. Perhaps this method can be used to describe better the relation between "simulated" and the "real" tasks.

An analogy or simulation should be readily interpretable by trainees. He or she needs to understand how the simulated objects/symbols represent the real objects. When the symbols are not understood (as in cultural differences) or are otherwise difficult to discriminate, their effectiveness is poor. This insures the salience for the learner. Familiarity with the representational system is thus an important determiner of the simulation's adequacy. It is probable that this factor is the basis for requiring different types of representations for people completely unfamiliar with the task and system to be learned than for those needing refresher training or broadening of their experience.

Another important characteristic is "transparency." The learner should be able to readily "see" the results of his actions on the system represented. In this sense, the representation may be quite different than the actual task. The simulation should reveal the interaction of the familiar visible entities represented. Table 2 summarizes these design guidelines.

Let's return to the example used for critiquing the state-of-the-art in the design of an ASW simulator. The example pointed out that the simulation was "high fidelity" in the naive physical sense. The equipment and the signals to be seen were realistic. However, the training on critical tasks received low emphasis, performance and feedback about errors was left up to the instructors. What characteristics should be included to design or redesign the simulator to make it an effective training device?

Fidelity of representation. Dynamic, hi-fidelity signal representation and including time dependent information about targets. Time can be varied to permit more effective instruction.

TABLE 2

DESIGN GUIDELINES

- **NOVICE OR JOURNEYMAN LEVEL?**
- **REPRESENTATION QUALITY**
 - **TRANSPARENT**
 - **UNDERSTANDABLE & FAMILIAR**
 - **PROVIDE QUALITATIVE UNDERSTANDING**
- **PROVIDE PRACTICE**
 - **CUES AND OPPORTUNITY FOR ERROR – DIAGNOSTIC**
- **ARRANGING TRAINING SEQUENCE**
 - **RELAX CERTAIN CONSTRAINTS LIKE TIME
ESPECIALLY EARLY TO PRODUCE BETTER UNDER-
STANDING**
- **SYSTEMATIC MEASUREMENT OF PERFORMANCE**
 - **DEFINED CONDITIONS AND PERFORMANCE
STANDARDS**

Analyze and identify the critical aspects of the task. From the analysis, specify the conditions and standard of performance measures to be used to assess student progress and for evaluating the training. Provide opportunity for making errors, especially common ones, and correct the reasons for them. Diagnose problems in student performance and understanding of task.

Vary the time for problem development appropriate to a student's level of learning. Slow things down for early learning or to allow the student to get used to time sharing among several signals or tasks. Change difficulty systematically and progressively to ensure an appropriate range of experience. Adjust conditions appropriately to maximize practice.

Reduce load on instructors by automating presentation of problems and performance measurement. Incorporate wide range of signals to provide appropriate experiences needed. Provide an expandable library of signals.

Any system should include methods for keeping track of student progress and providing informative, diagnostic feedback to him/her about performance tests. In addition, instructors and managers need information about student progress and about the adequacy of the course.

This brief listing of what should be included in the simulator(s) to make it an effective trainer needs to be operationalized before they can serve as design guides. Substantial progress is possible using available knowledge. The inclusion of these characteristics in simulators will do much to make them effective training devices and provide a potential means of assessment of competency.

CONCLUSION

This paper began by asking: "Is simulation fidelity the question?" In the sense of a naive realism, it is not. Realism may be useful for some things at some times, but it may not be providing the best training situation in many instances. Requiring it may actually slow learning down considerably. The characteristics of simulation needed for effective and efficient training are often not present in physically realistic situations. That is one thing that makes some tasks hard to learn.

It is easy to assume that a realistic training device will be an effective trainer. Such misapprehensions are difficult to overcome. When this assumption is made, the problem is that the training/teaching is approached in too general a fashion. Training requirements and objectives are unspecified and the performance criteria needed to judge the success or failure of trainees are therefore unavailable.

To attain training task fidelity in a simulation, it is necessary to describe the task precisely in terms of the way it needs to be represented mentally, and to provide conditions for the learning of the representation. These requirements are not met simply by providing high similarity to the actual job conditions. Testing the adequacy of people's performance, diagnosing problems of understanding that lead to errors, and systematic exposure or practice on infrequent, or dangerous, critical events, are needed. The real problem is to provide usable requirements for the development of simulation that can be included as design specifications. Fidelity requirements are derived directly from specific training objectives and conditions known to produce effective training. They cannot be specified solely in a general fashion based on the physical characteristics of the equipment and tasks to be accomplished.

Perhaps a more substantial problem is to convince the personnel and training system managers, and those who develop new weapons systems to include the guidelines in their specifications for the design of training simulation. For this, a political and educational effort is needed. To be effective, we need to be able to present the case for including training characteristics strongly enough to have them written into specification standards. We do have such standards for instructional program development (see Branson, et al, 1975 and deKleer & Brown, 1980).

There is one other problem that impedes progress in the development of specifications and standards for training simulator development. It is resistance to being evaluated. Often, users are encountered who are threatened by systematic measurement of performance. Their perception of the potential for punishment when simulators can measure skill levels leads them to avoid their use. This creates problems in monitoring the use of simulators and perpetuates a mental model of simulation that slows or prevents the adoption of more effective principles of design. Therefore, systematic implementation and management is needed that includes educating the users to overcome this bias.

RECOMMENDATION

Formal development of instructional prescriptions for task fidelity needs to be undertaken. This can be done on the basis of existing knowledge from research and practice. These prescriptions should then be included in the specification of all

training devices. Finally, an implementation and monitoring system should be formalized to insure the proper use of the training devices and to provide performance assessment. Training for instructors and training managers should include a review of the need for systematic performance measurement to assess the adequacy of training simulation.

REFERENCES

- Anderson, J.R., Greeno, J.G., Kline, P.J., & Neves, D.M. Acquisition of Problem-Solving Skill. In Anderson, J.R., (Ed.) Cognitive Skills and Their Acquisition. Hillsdale, N.J.: Erlbaum Associates, 1981, 191-230.
- Bott, R., A Study of Complex Learning, Theory and Methodologies. Unpublished doctoral dissertation, University of California, San Diego, 1978.
- Branson, R.K., Rayner, G.T., Cox, J.G., Furman, J.P., King, F.J., & Harnum, W.H. Interservice Procedures for Instructional Systems Development. (NAVEDTRA 106A) Pensacola, FL: Navy Education and Training Command, 1975.
- de Kleer, J. & Brown, J.S. Mental Models of Physical Mechanisms and Their Acquisition. In Anderson, J.R., (Ed.) Cognitive Skills and Their Acquisition. Hillsdale, N.J.: L. Erlbaum Associates, 1981, 285-310.
- de Kleer, J. & Brown, J.S., Mental Models of Physical Mechanisms. Palo Alto, CA: Xerox Palo Alto Research Center, Paper dated December 6, 1980.
- CNETINST 1550.15, Chief of Naval Education and Training, 29 May 1980.
- Gentner, D. The Structure of Analogical Models in Science. (Report No. 4451) Cambridge, MA: Bolt, Beranek, and Newman, Inc., July 1980.
- Hollan, J. & Williams, M. Human Reasoning About a Simple Physical System: A First Pass. Paper in preparation, 1981.
- Johnson, S.L. Effect of Training Device on Retention and Transfer of a Procedural Task. Human Factors, 1981, 23 (3), 257-272.
- Johnson-Laird, P.N. Mental Models in Cognitive Science. Cognitive Science, 1981, 4, 71-115.
- Larkin, J. The Role of Problem Representation in Physics. Paper presented at the Mental Models Conference, University of California, San Diego, October 1980.
- Mackie, R.R. & Shultz, T.E. An Analysis of the Potential Use of the Sonar Performance Testing System (SPTS) for Training Critical Operator Skills. San Diego, CA: Navy Personnel Research and Development Center, Special Report SR-xx, submitted May 1981.

Montague, W.E., Ellis, J.A., & Wulfeck, W.H. After Years of Instructional Research Do We Know More Than Grandma Did About How to Teach People? Invited Address, Annual Meeting of the American Educational Research Association, Session 23.02, 15 April 1981. (To appear in ERIC).

Riley, Representations and the Acquisition of Problem-Solving Skill in Basic Electricity/Electronics. Paper based on a talk given at the Computer-Based Instructional Systems and Simulation meeting sponsored by the Office of Naval Research, Personnel and Training Research Programs, Carnegie-Mellon University, January 15-16, 1981.

Rumelhart, D.E., & Norman, D.A. Analogical Processes in Learning. (CHIP Report No. 97) La Jolla, CA: Center for Human Information Processing, University of California, San Diego, October, 1980.

Semple, C.A., Hennesy, R.T., Sanders, M.S., Cross, B.K., Beith, B.H., McCauley, M.E., Aircrew Training Devices: Fidelity Features. Wright-Patterson AFB, OH: Logistics and Training Division, AFHRL-TR-80-36, January, 1981. Final Report-Aircrew Training Devices: Fidelity Features.

SIMULATOR USE IN FOUR FUNCTIONAL AREAS

Leon H. Nawrocki
U.S. Army Research Institute

REPRODUCING PAGE BLANK-NOT FILLED

ABSTRACT

The military use of simulation is discussed in 4 functional areas: selection, training, sustainment, and evaluation. One must decide on the orientation for simulation then decide what the relevant fidelity issues are within that functional area. A discussion of two groups who are against high fidelity and total realism is presented.

For the last three or four months I have had several discussion with people in the ARI Training Lab who have thoughts on the area of fidelity. My own opinion has been that fidelity of simulation is a nonissue. Concurrently, I have been trying to put together, and I say trying because it has been difficult, an overview paper on game simulation with an emphasis on small electronic games. Arcade and personal electronic games are a big fad these days, and in examining the topic, I had, of course, to address the whole area of simulation per se. What you are going to see is a scaled-down version of what I have been thinking about in the context of the game simulation paper. Scaled-down in the sense I am going to focus on simulation. I have two points to make. The next two tables address point one:

(Table 1)

I've spent two or three weeks over umpty-umpty definitions of games and simulations and I think it is worthwhile clarifying what we mean by simulation because you find for each person there is different definitions. I prefer the top definition: an operating representation of events and processes. For one thing, it is simple. For another thing, it does not constrain you to the other use of simulation that people really haven't addressed, although Jack Thorpe alluded to it, which is simulation where we are just modeling a system. We are not really worried about human interaction in this latter case. In the military, for example, you may be trying to model the world to test doctrine. I believe this group is probably interested in the people part and the inputting; i.e., what happens when we do something to this simulated world. So I am going to focus on using the first definition, simple, straightforward.

Given that orientation, and remember that word orientation, one of the things I did in looking at games was to say, "Gee, what do we do in the military that covers the whole training system?"

(Table 2)

Well, primarily there are only four major functions. You can argue about the terms I use, but first we pick people. We select them for whatever reasons, either from scratch or because they are already in and we want to retrain them. We then initially train them or bring them into an instructional segment. What we are concerned with after that is proficiency maintenance or keeping them going--practice. The Army uses sustainment because otherwise they get confused with the word maintenance in terms of actually maintaining things. That is what I was told. So sustainment refers to practice or maintenance or proficiency of skills. The last thing is that periodically we want to evaluate people for a number of reasons: promotion, training adequacy, readiness, etc. These then are the four major kinds of functions we are talking about. What happens when we use simulators to support these particular functions is that there are issues that aren't really

TABLE 1

ORIENTATION	SIMULATION: TWO VIEWS
PEOPLE AND INPUT	AN OPERATING REPRESENTATION OF EVENTS AND PROCESSES
PROCESS AND OUTCOME	A TECHNIQUE USED TO STUDY AND ANALYZE THE OPERATION AND BEHAVIOR, BY MEANS OF SYSTEMS CONDITIONED BY HUMAN DECISION AND/OR PROBABILISTIC NATURAL INFLUENCE

TABLE 2

PEOPLE FOCUS AND "FIDELITY" ISSUES

<u>SIMULATION FUNCTION</u>	<u>"FIDELITY" ISSUE</u>
SELECTION	SKILL VS TASK
TRAINING	PART VS WHOLE
SUSTAINMENT	EXPERIENCE VS TRAINING
EVALUATION	NORMATIVE VS CRITERION

fidelity issues. That's why the word fidelity is in quotes, and is my personal opinion.

I have listed only one issue for each of these functions and the "versus" may be a bit strong. My point is sort of a follow-on to Montague's that we may be focusing on the wrong issue if we just start running around talking about fidelity. In the case of selection, I would argue that we would probably want high fidelity in terms of measuring the skills that our selection is based upon, but we are probably not terribly concerned about task fidelity at this stage, particularly if we are talking about a new recruit. The recruit doesn't know the job, so it is not really fair to test him in a high fidelity task. So the issue here may be the degree to which we have fidelity of the skills that we are trying to measure by aptitude. There are other issues, but I am mentioning what I consider key ones.

In terms of initial training, I think the real issue here is part versus whole, which is an issue that has been with us, like the Holy Grail, for something like 30 or 40 years, but that is the real key in training. There is a fair amount of evidence that if we have a highly realistic simulation or simulator in initial training, we may, in fact, get negative transfer so we are really talking about the extent we are trying to do part training. All we want is fidelity in a particular part task and we don't need the whole shebang, if you will.

When we are talking about sustainment, the person is already on the job and there are two things we are really concerned about. Are we using a simulator in the sense of training, like retaining or refreshing, or are we really concerned with providing the person with the kind of experience we expect them to encounter in a real world? That's a nontrivial issue. There is a fairly rational argument that one of the reasons one goes out in field exercises is to get the feel of what it is like trying to do your job when you are rumbling around in the dirt and grime. That may not really be training in the sense we think of training or instruction, but it has a useful experiential value to it, so there is not the shock of a new world out there should the balloon go up. So what one ought to consider here is: are you looking for experiential practice where you might want to go for fairly high realism, or are you really concerned about training in which case you are back to the first issue and may want reduced fidelity.

Lastly, in terms of evaluation, one of the things we should consider is evaluating people on a normative or a criterion-referenced basis. If you are talking about normative, what you are really interested in is a situation where the simulator that you are using doesn't need much flexibility. It can be a fairly fixed scenario because you want to compare all people according to the same thing. In the case of criterion referencing, what you want is probably a more realistic simulator that has a great deal of fidelity and a lot of flexibility. That

is the key here. So the real issue here is degree of flexibility. My point is that for workshop purposes one ought to say, "Okay, what is our orientation for simulation? Is it training (and my assumption is, it is) or just to try out processes per se?" Once you have decided that, consider the function for that simulator, and I have provided one view of the world with four possible functions. I think once you have done that you then ask what is the relevant issue within that functional area?

I have some additional comments based on listening to the previous speakers, which I think may be of use. Rather than coming from my paper that I've been working on and keep talking about, these comments are in the context of my role down at Training and Doctrine Command, where, of course, I deal with a number of users (and developers for that matter). What kind of questions do you get that relate to fidelity? Basically, you only get one. The question is always, "Do I really need to have function X in this simulator?" The answer to that question is not a fidelity question. The answer to that question is, Does X provide that particular function that you want to train? For example, the issue may not be whether or not you need color because it provide greater fidelity. The real issue more likely is does color provides the person with the necessary knowledge needed to perform that job. That's a separate question. So again I am saying, think about the relevant issues.

Two additional observations. Who in fact, seems to be against high fidelity and total realism? Well, there are only two types of folks and two reasons. The first are some of the bright people like ourselves who say, "Gee, you may actually get negative transfer, you can interfere with what you are trying to teach." The second group are the people concerned about dollars and their argument is always, can we afford it? So these are the only two negatives you will ever typically hear. On the other hand, those who favor high fidelity are the users, primarily because they say, "It looks right, feels right, must be right." The second group of people who sometimes favor high fidelity are us bright folks again who don't have the answers. So what we typically do when we're not sure about level of fidelity is we simply shotgun it. Because you figure, oh well, everything is better than missing something. There is another group that often supports high fidelity and that is the people who make money by building simulators. We can't ignore that, because they are going to say, "Let's have more realism."

So there are at least two sets of low fidelity arguments: costs too much, and the possibility of negative transfer; several high fidelity arguments. From that, I would argue that the person who is buying the simulator is the crucial person here, in the sense of the buyer/user, and I think what needs to be done is in terms of processing and monitoring the purchasing of simulators (I have been arguing this point for years). It ought to be our job to tell somebody, to the extent we can, whether a function or capability is going to provide them what they want, in terms of

their goals, or whether it is simply there for reasons of user acceptance. If we make that clear and if the user has the extra dollars, fine. There is absolutely nothing wrong with user acceptance, but the cost of that then becomes a decision maker's problem, not a technician's problem. I would also argue, for us who have to live in a bureaucracy, that such decisions ought to be documented somewhere. Because what eventually happens, as some of you are well aware, is once you get a high-speed, heavy duty, dual, all-purpose simulator out there and somebody begins to look back at the dollar cost, that somebody always comes back to the technical developer and says, "Why did you add the blip?" In many instances, as we all know, it was added in because the user said, "I won't buy that thing unless it looks like this." Fine, a reasonable argument that should be documented. I can't do a thing about that, but that's something people ought to think about. Those are the only points that I wanted to make and I hope again that it helps somewhat in the people focusing during the workshop on some of the "real issues".

SUMMARIES OF TOPIC AREAS IN THE WORKING GROUP SESSIONS

Garvin L. Holman
Army Research Institute
Thomas J. Houston
Army Research Institute
Gary A. Klein
Klein Associates
Robert W. Swezey
Science Applications, Inc.

ABSTRACT

Summaries of the four topic areas discussed in the working groups are presented. The topic areas were:

- (1) Effectiveness
- (2) Fidelity
- (3) Guidance
- (4) Priorities and support

Summary

EFFECTIVENESS SESSION

Garvin L. Holman
US Army Research Institute

It is a little difficult to summarize everything that went on in those several hours of work group session. As you might well guess, we heard a lot of war stories, a lot of anecdotes and a lot of fairly sophisticated comments on the problems we were discussing which probably didn't get recorded and should have, and probably a lot of trivia that did get recorded. I take it, as my task here, to give you some sort of synthesis, or at least a summary, of some of the discussion that went on in the Simulator Training Evaluation section.

We concluded, after several hours of discussion, that if you are going to evaluate a training device, the way to do it is with a transfer of training experiment. Now this probably comes as no great surprise to anyone and we only had a few people who were dissident to that point of view, but they were quickly beaten into submission. By transfer of training experiment we meant the procedure by which a group of trainees is trained to perform certain tasks in a simulator and then tested/re-trained on the operational equipment. This group's performance on the operational equipment is compared to another group's performance that was exposed to an alternative training system (usually training on the operational equipment alone) as an indication of the training effectiveness of the simulator. The transfer of training experiment is the state of the art for determining the training effectiveness of a simulator or training device. It makes little difference in principle, I suppose, what sort of measures you end up with: transfer effectiveness ratios, first trial proficiency, some other savings measure, or whatever is most appropriate to the system that you are evaluating. Whatever your goals are, you are going to select the appropriate measure. We also concluded that it is difficult to perform a perfect transfer of training experiment since the world does not hold still, much less so in the field environment than it does in the laboratory, but you can give it a good shot. All but a few of the transfer of training reports in the literature have been victims of too limited resources, uncontrollable variance or serious confounding, but those few demonstrate that these experiments can be done. If a service is serious about evaluating the effectiveness of a simulator, it should be willing to make available the resources required to conduct a good transfer of training experiment.

The transfer of training experiment was contrasted with other methods of evaluating the effectiveness of a simulator such as

instructor and student opinion. This is a widely used method, but a method widely criticized for very good reasons which we need not go into here. Improved performance on the simulator itself is sometimes used as an evaluation technique but we concluded that it only shows that the students learned something on the simulator. Whether what was learned was what the student needed to know to perform well on the operational equipment is not indicated. Another technique is to generalize from other situations, from what can be tested in the laboratory or the field. Again, these generalization methods are not considered satisfactory. There are also analytic methods that could be used to estimate training effectiveness such as the Army's TRAINVICE models and the Air Force's Comparability Analysis. These methods suffer from being analytical rather than empirical. However, ARI is currently working on these methods for prescriptive and predictive purposes.

The major question that was raised in this session dealing with effectiveness was one that was brought up in yesterday morning's session: What research needs to be performed to assure us (and us is a lot of people, scientists, training systems developers, simulator manufacturers, trainers, equipment operators and maintainers) that tasks or skills that are trained on a simulator, that cannot be performed on the operational equipment, are being properly learned on the simulator and will be transferred effectively to the operational equipment? Examples of these tasks that may be trained using a simulator but not performed on the operational equipment are emergency procedures or combat skills and tasks that are too dangerous to perform in a non-emergency or non-combat situation. This means that those tasks cannot be trained on the operational equipment and therefore are not subject to transfer of training testing that requires validation of the simulator training by re-training on the operational equipment. The first blush answer was, "We can't." Well, we weren't looking for answers like that. At the very least we were looking for research questions asking how we can go about researching that problem in the absence of the transfer of training experiment. Several research issues came to mind. I will list them with some explanation of perhaps what was meant.

A question was asked, what skills are sensitive to changes in the state of the trainee, such as stress, that may differ from the simulator to the operational environment? How can training be improved to overcome these effects? The classic example is the fighter pilot being shot at. He can be shot at in the simulator and that's not near as stressful as being shot at in the real world. Does this make a difference in his performance? How should we train them to make up for that? No answers, just questions.

More in the line of an answer or a suggestion of how we can assure ourselves that these nontestable-on-the-operational-equipment skills are properly trained went like this. Careful training development and simulation development is the answer to assure us that all is right. If the simulator is developed and

manufactured properly we need not worry about it. If it is designed to train these skills and built to train these skills, then it will train these skills. This ignores the fact that most simulators are developed and built as well as is known how and they still do not teach all the tasks they were designed to teach as well as they should. The research questions that pop out of that comment, however, deal with front end analysis. Such as, was a good front end analysis done? What do you mean by good front end analysis? What constitutes a proper task analysis? What constitutes a proper simulator specification so that the manufacturer knows what to construct? How do you know your training program and its use are the appropriate ones to train those skills. Now those are very general questions but they are questions for research. A lot of them are fidelity type questions that have probably been gone over in other sessions. The general suggestion was that research is needed to determine the principles of simulator development by which we would believe that the situation is all right, that those untestable skills are being properly taught. That's a big question.

Another suggested piece of research dealt with what we came to call control strategies. That is, there is a perceived need to determine that the responses being made in a simulator are the responses that should be made at the appropriate time in the operational equipment. One should attempt to do this before the simulator is developed. To design the simulator with that in mind would be part of the task analysis at the start and it could also be part of an operational test. One should check and see whether or not the control strategies or responses that are made in the simulator in these nontestable tasks are the ones that are believed to be necessary in the operational equipment. That would be a form of training effectiveness test that one might do in the simulator. It is suggested that research needs to be done to validate that notion and to develop research strategies so that this kind of test procedure could be done routinely where appropriate.

Another suggestion for research was to study failures in simulator training. Every simulator is notably not a perfect simulator. There are tasks and skills that any simulator was presumably designed to train, that it does not train as well as we would like. It was pointed out in the sessions that very rarely were these instances researched. Very rarely do people do the research necessary to find out why a presumably well developed simulator isn't teaching what it was developed to teach. We saw a need for surveying simulators for common failures to train. There may be clues in that as to what is going wrong. Also, research is required on the individual simulators to determine what they do wrong. What is the problem either in our analysis starting at the very beginning, our development cycle, or our technology that makes certain skills poorly trained on a simulator?

Another item dealt with performance measurement. Research is needed on performance measurement to insure that instructors look

at the relevant behaviors of their students. An old problem but it is still sticking its head up. You can't give proper feedback if you don't know what to look for. The hope here is that if the performance measurement is correct on all of the skills you can test in a transfer of training experiment, there is a good chance they are correct on those that you can't. And consequently you can be more assured that those skills are properly trained.

The last item that I will mention is the need to do some research on simulator instructors. We need to know how they perform now, which many people pointed out in some cases was poor, and how they ought to perform. And programs need to be designed to teach simulator instructors to teach in their simulators, in order to turn out the best student that they can.

Those are a list of research issues or research questions that were brought out. We didn't get to the part of the agenda where we should have been talking about methodologies and so on. The methodologies are probably a matter of the individuals conducting the research to sort out and to implement. We did not get to the question of who ought to be doing this research. Hopefully one of the other groups has been talking about things like that. So I would like to conclude with that remark.

Summary

FIDELITY SESSION

Thomas J. Houston
US Army Research Institute

The topic to be discussed in this portion of the simulator fidelity workshop was fidelity. There were two general areas to be covered. (1) Measurement of Fidelity, and (2) Generalizability of flight simulation data to other areas of simulation training. Under the first topic were such questions as: What is fidelity? Is it measurable? If so, how and why and what relationship is there if any, between fidelity and training? The second topic resolved itself to questions of feasibility and desirability of generalizing from flight simulation data to other areas of simulation, given that most of the available data is based on flight simulation. Some how none of the groups managed to address these two issues separately, but there was some information generated on each one.

The question of fidelity: We started off each session generally with the general question, Should we talk about fidelity at all? Some people said yes and some said no. There was quite a bit of feeling expressed at some length, that perhaps training effectiveness was the more central issue. No one disagreed with that statement, but we were not there to talk training effectiveness. We were talking fidelity. We did arrive at an attempt at a definition of fidelity. The feelings about fidelity generally ranged from fidelity being a nonissue to fidelity being the most important thing in simulation. The concensus, arrived at independently by all groups was that if we are going to talk about fidelity as a global entity we are in trouble. There are aspects of fidelity which can be dealt with individually or in conjunction with each other, depending on the particular situation which we face. There is fidelity depending on the simulation that you are attempting to do. There is fidelity depending on other things. This got us into a discussion on what types of simulation we generally deal with and from that emerged most of the work on the transfer of information from flight simulation data to other types of simulation.

There were several definitions of fidelity offered. There was a serious attempt to see if we could find a definition which is task independent. There was no agreement reached on that. I suspect the reason was that we were going at it this time with the global issue of fidelity and that didn't seem to get us very far. We more or less agreed that the term fidelity, regardless of what definition we are talking about, carries the implication of physical fidelity, especially to users and some of the hardware

vendors. We decided that it really should not do that and that the types of fidelity that we should be talking about, the aspects of fidelity that we should be talking about, are such things as procedural fidelity, and an interesting one, fidelity by subterfuge. I will get into some definitions of these in a minute.

Physical fidelity, I think we all agree, is the extent to which our device, our simulator, resembles the actual equipment. Task fidelity we defined as the extent to which the tasks that we are attempting to train are trained. Perceptual and phenomenal fidelity were defined in terms of the extent to which the simulator or the training device applies itself to the human system receivers to which it is directed. Examples given were in terms of range of human hearing, minutes of arc which a human being can observe and things of that nature. There was a bit of discussion on that. There were also some discussions in the relatively esoteric area of cognitive fidelity brought on by our representative from Ft. Leavenworth attempting to simulate an entire battlefield and we all decided that was impossible. There was a distinction made, apparently, between the continuous tracking type of job such as in flight simulator, tank driver, or operations in various fields and procedural type tasks such as in maintenance trainers and also between the various aspects of fidelity applied in various different way to these two tasks, these two types of simulation. There was quite a bit of discussion concerning a metric for fidelity. The upshot of all this was that we should probably not attempt to measure fidelity on a linear scale but rather a branch-and-flowchart identifying the categories and procedures of various aspects of fidelity which apply to the situation. It was felt that this would enable the training developers, the training systems people to speak more adequately to the engineers in the development of what types of fidelity were necessary to build into the simulators.

There was a bit of discussion on fidelity relative to the behavior. And that brought about the idea that fidelity is not necessarily a representation, even in task fidelity, of the actual task. That there are many instances in which a representation of the actual task is not a good way, far from being the best way, not a good way at all, to teach the task and that deliberate degradation of fidelity from the task is necessary in order to promote good teaching. That was about the extent of the fidelity issues that were covered.

The subject of transfer of fidelity data from flight simulation studies arose, as I mentioned before, from the difference between the continuous tracking operations and the procedural. There seemed to be no agreement on methods and procedures from transferring such data but there was agreement that current research seems to be headed in that direction and that possibly flight simulation data could be used in the construction of simulators for such things as tank drivers and vehicle operators. One got the distinct feeling that given

another week of intense discussion we could have achieved a major breakthrough. We'll have to settle, instead, for a few years of research, nudged in the breakthrough direction by our discussion.

Summary

GUIDANCE SESSION

Gary A. Klein
Klein Associates

The question that we were dealing with was how to provide guidance for fidelity decisions. Should we build training devices with high or low fidelity?

The context that we were working in was how to buy training devices that would produce more effective training. We made the assumption that fidelity research and guidelines were available. We assumed that they would be coming out of session two and therefore in session three we had to figure out how to use those data. Who is going to use them in what form? What type of guidance were we going to have, and who was going to provide that guidance?

I am going to summarize the general reactions of the group. Not everybody in each group agreed with all these things and there may be some strong disagreements but there seemed to be a consensus.

The general consensus seemed to be that people wanted to reject the concept of fidelity within that context for that purpose. The fidelity question - should we build a trainer with high or low fidelity - did not seem to be providing much useful information and it might be a question that it was going to mislead more than assist. We decided to reject the concept of fidelity as a useful concept for three reasons.

(a) First of all, because of the problem of emphasis. There is a general class of design decisions that people have to make when they are designing a training device. Focusing on fidelity may create the wrong emphasis. People really have to figure out who is going to be trained, for what, what type of training program, what type of environment, what does the task consist of, how will we determine when the task has been adequately trained. If you deal with those issues, then the fidelity question seems to be taken care of. But if you begin by focusing on fidelity, then you are focusing on engineering issues not on training design issues. You may be ignoring training design issues. So it seems to be that there is a potential for misplacing the emphasis.

We were not saying that the fidelity question was illegitimate, just that it may not be very useful. It may not be helpful to ask whether a device should have high, medium or low fidelity without putting this in the context of training

requirements; but once you start examining training requirements, you may not need to be concerned with the general question of whether the device should have high, medium or low fidelity.

(b) The second reason for wanting to reject the fidelity question as useful, within the context, was a theoretical one. Consider the following example. Let's say your task is to train somebody to ride a motorcycle and you can have two possible training devices, one is a bicycle and the other is a Volkswagen convertible. You might ask the question about physical fidelity. Does the bicycle or the Volkswagen convertible have more physical fidelity to the motorcycle? And you go on forever because each of them have an infinite number of physical features in common. You can do this in a philosophical sense by saying that both of them are closer to the earth than the moon and the moon than one star and one star than another and you can go on for the infinite number of stars and you can say that each one has an infinite number of features in common. So if you are going to define fidelity in terms of common features you are not going to get to anything that's realistic. Similarly from functional fidelity, you can think of an infinite number of ways in which they can be used similarly and not arrive at any useful end point. What you need to be doing is asking, are these two things similar with regards to some purpose, some task, some goal? Of these two potential training devices, which one is going to have the right amount of similarity to train people to do a certain task. But then you are asking the training requirements question and you may not need the question of fidelity. How are you going to specify the task to be trained and how are we going to use that specification to give guidance to the engineers so they can build a responsive machine?

(c) The third reason for wanting to reject the concept of fidelity was a temporal one. Most training devices require changes after a short period of time. A device may have high fidelity at one point, but two months later, six months later, it will have less fidelity, because the actual equipment will be modified. So fidelity seems to be a very fleeting phenomenon.

Since we wanted to replace the question of fidelity, we needed to identify some better questions and this process continued over each of the sessions that were conducted. We considered this making progress because it seemed that a useful way to make progress is to start with reasonably good questions, find ways in which they can be improved and hopefully wind up by asking better questions. We were hopefully making progress by finding better questions to ask as a result of this type of analysis. Instead of talking about fidelity, we decided to discuss the process of designing to training goals by asking what is the nature of the tasks to be trained, what type of personnel are going to be trained, to what level, what kind of training context and special practices are going to be employed. (These are also important with regard to instructional features.) All of these decisions need to be made. All of these issues need to be resolved in order

for engineers to figure out how to design the machine. Once these issues are resolved, then designers have answered what we would formerly call fidelity questions.

What do we need in order to design to training goals? Needs mostly are in the area of communication. We have to communicate things to engineers that we haven't been able to communicate before with regard to mental models or sensory/motor skills or perceptual learning, or cognitive tasks. We don't know how to represent these. We don't know how to examine these skills, except possibly in very laborious ways that may not even work for all kinds of tasks. So we need to have a much better way to represent to ourselves what the nature of the task is and then to communicate that to the engineer. As long as that's missing, then we can substitute general questions about how much fidelity is needed, but we're really responding to a lack of good ways for structuring our task description methods. We just don't know how to define complex skills and formats that appear to encompass the needed wisdom without actually presenting it.

We came up with several recommendations about what could be done to move us in a better direction and provide necessary guidance. (a) First of all, we thought it would be useful to have historical studies. These would be lessons learned on existing simulators. The reasons for decisions, reasons why certain devices work, and other do not, attempts to trace failures or successes to specific features of training practices. (b) Second, we thought it might be useful to develop handbooks, or make available handbooks, such as the Integrated Perceptual Information for Designers Handbook currently being developed by AMRL. Possibly that information might provide useful guidance for engineers, psychologists and training designers. (c) Third, we were worried about the issue of guidance if that guidance was going to consist of technical aphorisms that were filed away, forgotten or misinterpreted. We thought that corporate memory was more important than guidance. Corporate memory is a living guidance. We must recognize that we can't write things down cleverly enough to force naive people to make smart decisions. We have to get the personnel experienced so they understand what the issues are, how to interpret the data, what kind of factors to be sensitive to. People familiar with the training devices of 20 years ago need to try out the newest operational models in order to make good decisions. (d) There was a concern expressed, in a number of our sessions, that the reinforcements are all wrong. The person procuring or building the device is reinforced for getting the device out under budget and on time and never has to have any concern for whether it works in the field, whether it trains, whether it does what it is supposed to. This reinforcement system does not seem to be an optimal one for motivating people to build effective devices. Hopefully there could be other organizational systems in which personnel might design a device, and then go out and do some initial training with the device so that people in the field could see how it should be used, and then cycle back to design another device. This raises

the whole issue of organizational changes that might be necessary in order to maintain and develop expertise within an organization, and allow corporate experience to emerge. We saw that as happening to some extent. We saw that people were becoming more sophisticated about training devices. Personnel moving up into higher command levels were getting experience with better simulators. We thought that this process possibly could be improved by modifying organizational formats but we didn't want to spend too much time making recommendations in this area because of the difficulties of reorganizing. (e) A fifth concept that arose was one that was introduced by one of the speakers yesterday. We decided to come out in favor of back-end analysis or proctological analysis, as opposed to front-end analysis. The reasoning here is that there is a lot of concern paid to how you design, select its features, determine what the device is supposed to do. Maybe a useful reemphasis would be the back end. How is the device going to be evaluated? Who is going to decide whether it is successful or not? How are you going to make the decision that the simulator is doing what it is supposed to, that the level of training is what was intended and expected and needed? There is some discussion of criteria at the beginning, but criteria do not seem to get connected to design questions. You still have to get into an analysis of the task whether you start from the front or the back. An advantage of starting from the back is that you tend not to forget or de-emphasize the criteria issues. A back-end analysis would force people to pay attention to very critical issues of how you are going to evaluate whether your device is going to be successful and how you use that information to affect your design of your device. (f) Another possibility for providing guidance would be to provide examples of good inexpensive training devices, possibly developed by laboratories. This would provide a wider set of analogues that people can observe and use to figure out what they can do to train new tasks and new devices. (g) Another suggestion would be the development of analytical techniques to define simulator training needs for engineers. We are not sure that ISD methods are going to work. They don't seem to have worked very effectively so far. Maybe they can be modified, maybe not, but this seems to be a controversy. There may be a need to develop a better approach to defining tasks so that psychologists can get into the critical issue of what needs to be trained and communicate these training needs to engineers. Here we get into a dilemma. Consider two alternative. One is to define for the engineers how many corners and edges need to be put into the device for effective training. The engineers would appreciate this type of description but we are not sure that we can provide it. Secondly, we could simply tell the engineers that we expect 50% transfer of training on certain tasks, and we could do that, but the engineers would resist it because we would be just pushing the training design decision onto the engineers. These two alternatives are two sides of the same coin. We need to be able to define the task well enough to talk to the engineers. Once we have a clear understanding of what perceptual discriminations the trainee needs to be able to make, then we have the basis for talking about edges and corners. Once we have

understood the task well enough to be able to say this is what the person ought to be able to do, then we have a basis for talking about a transfer of training evaluation. There doesn't seem to be any way of avoiding the need for clarifying the task descriptions. That is our conclusion about the type of guidance needed in order to make fidelity decisions.

A second question that our sessions dealt with was the introduction of new technologies. There seemed to be a general agreement that new technologies were often introduced but were not always used effectively. There was a sharing of horror stories about features that were never turned on, about devices that were never turned on, about features that were used inappropriately. There seemed to be a general concern that new technology is diffused effectively. Examples were automatic performance measurement, freeze, record and display, videodisc. Some suggestions emerged. (a) First, we thought that part of the problem was a lack of communication between the designers and the users. If these new technologies are to be coming out, the people who are going to train with them have to be part of the process, so they can learn and advise. It may not make sense to develop a playback capability where you can play back the last seven minutes. If I only need the last 30 seconds and I have to wait six and a half minutes to get it, I am not going to play it back. Or if you give 50 pages of hard copy and I only wanted to look at two summary statistics, I am going to throw the whole thing out. So it seems advisable to get your users involved in the new technology development. (b) Second, new technology will almost always have bugs. Nothing ever comes out right the first time. Hopefully, there will be time to get these bugs worked out in the laboratories and this could be part of the laboratory mission. To take new technologies, do demonstrations and see just what works, what doesn't, what needs to be modified, where the shortcomings are, what the boundary conditions are. (c) Third, we felt there were many valid cases of rejection of new technologies. Until the bugs were worked out, training personnel were making a wise decision not to use the features because they were just going to reduce the quality of the training program. (d) Fourth, there was a concern about human-machine interfaces and the necessary human factors engineering. These costs are usually avoided to keep the device costs down and then, if necessary, they are added later. But human factors engineering is more expensive if included at the end of the design process, and it may be too late to generate viable recommendations. So there was a suggestion that human factors engineering be added in at the beginning even if it drives the cost up a bit initially. (e) There was a feeling that maybe we should try to build in flexibility where possible in order to solve some of the problems about updating. (f) A final suggestion that emerged from the sessions is that of a liaison type of contact, a facilitator. The analogy was chosen of an agricultural extension agent. Somebody who lives in a farming community and is a conduit of new technologies to the local farmers. He can talk to the farmers and ask if they are aware of existing technologies. And they can make him aware of problems needing new technologies.

In the simulator world, once the device goes out, nobody follows it to see whether it is doing what it is supposed to. Are the users turning on all the switches? Are they using all the capabilities? Are they misusing the capabilities? It is rare that government personnel or contractors go out to check utilization. The device is delivered and just left there. The suggestion was that someone like an NCO, who is involved in maintaining and sometimes running the device, could take on that kind of special role as a conduit of technology at a given location. He/she could go back to the engineers, to the designers and tell them, here is the way the device is being used and here are the problems. And they can point out that the operators are not using the device in the way it was intended. Or problems with the device can be filtered back to the engineers or the training developers, to identify needs for retrofits, and to enable the generation of newer designs and approaches. So some sort of liaison contact could be a feasible way of ensuring communication between the people who are using the device and the designers once the device gets out into the field.

Summary

RESEARCH PRIORITIES AND SUPPORT SESSION

Robert W. Swezey
Science Applications, Incorporated

INTRODUCTION

This paper reports on discussions held at the U.S. Army Research Institute Workshop on Simulator Fidelity, 23-24 July 1981, concerning the topic session on Research Priorities and Support. The purpose of this session was to address two principal topic areas:

1. Identification of research issues of interest in the general areas of fidelity, simulation, and transfer of training, and
2. Given that research issues in these areas are both complex and long-range, identification of methods of generating necessary long-term support for conducting such research.

The discussion which follows summarizes the issues raised in this session of the Fidelity Workshop.

Research Issues

Research issues in the area of simulation and fidelity were classified into two broad categories according to their level of abstraction. On the one hand were issues termed "global," focusing on questions about the nature of simulation as a training medium, the appropriateness of the fidelity concept in evaluating training device effectiveness, etc. On the other hand were more concrete issues which were felt to be possibly conducive to laboratory analysis or experimentation. These issues dealt with the more specific topics of individual differences, motivation, performance measurement, etc., as they related to the design and evaluation of training devices. Within this broad dichotomy, several issues emerged which were considered to be major research priorities currently facing the training community. These were:

o Global Issues

- Is the concept of fidelity viable and appropriate; or should this concept be eliminated entirely in favor of research on training device effectiveness?

Fidelity appears to be a multifaceted concept, the nature of which is not clearly defined. There appears to be a general consensus that fidelity can be understood in terms of two primary referents: the device per se (physical fidelity) and the function simulated (perceptual, or phenomenal, fidelity). While this distinction is generally recognized, the actual role of fidelity in training device design and evaluation continues to be a hotly debated issue. At one extreme are those who advocate the position "the higher the fidelity, the better the training device." Engineers and manufacturers developing high-fidelity simulators often tend to embody this approach. Others follow a middle-of-the-road position, and point to the plethora of research findings which show that while in some cases high fidelity devices lead to high transfer of training, in others, low fidelity devices accomplish the same results. This position suggests a contingency approach to the level and nature of training device fidelity. Finally, at the other extreme, are those who believe that fidelity is not an issue in training device design. This position advocated a reorientation in research away from the fidelity question and towards the problem of developing methods to increase learning and transfer (i.e., develop devices which best facilitate transfer of training independent of the fidelity issue).

- How can communication between engineers and behavioral scientists regarding training devices be improved?

By the very nature of their professional concerns, engineers and behavioral scientists often tend to differ in orientation toward training devices. Engineers are typically concerned with hardware issues, whereas behavioral scientists are more typically concerned with the training potential of the device. These different viewpoints often result in a communication gap. Therefore, a need exists to develop a mechanism for facilitating communication between such individuals. Several possibilities were suggested toward facilitating this end, including: (a) helping behavioral scientists to be more aware of the constraints of costs and scheduling in manufacturing; (b) helping engineers and consumers to be equally concerned with the effectiveness of a training

device as with its fidelity; (c) developing better methods of quantifying performance criteria so that they can be incorporated into device specifications; and (d) creating new job skills which ultimately would lead toward a composite of engineering and behavioral scientist viewpoints.

o Issues Conducive to Laboratory Analysis

-- It was recommended that basic research be conducted on the following issues as they impact training device effectiveness:

--- Individual differences - There appears to be little research on the effect of individual differences on training device effectiveness. Research on this issue is, however, beginning to gain some support in educational psychology, as evidenced by increasing awareness that learning devices and methodologies are most successful when tailored to the particular cognitive style and capabilities of the learner.

--- Skill level variances among trainees and trainers - This issue is related to that of individual differences (above). Research is needed to address the relationship between trainee skill level (e.g., input repertoire) and various types of training devices. As skill levels vary across trainees, so perhaps should the type of device used to train a specific set of behaviors. Similarly, trainers vary in their teaching skills. Some devices may be more appropriate for certain types of trainers and conversely, various trainers should possibly receive differential amounts of instruction before utilizing advanced training devices. The whole issue of trainer-training has been slighted in research up to this point.

--- Basic learning processes. One important research issue is identification of aspects of the basic learning process which can be incorporated into training device design in order to maximize transfer, rate, and other characteristics of skill learning.

--- Performance measurement - A research issue involves developing methods for quantifying required student performance measures in

order that they may be incorporated into device design specifications rather than tacked on after devices have been completed. Inherent in this issue is the specification of feedback capabilities, interactive training capabilities, etc.

- User acceptance - A training device is effective only to the extent that it is used. Because operational environments in which training occurs are often less than ideal, it is important that devices be perceived as useful. Research is therefore needed to determine ways of maximizing the acceptability of training devices.

- How best can transfer of training and training effectiveness be measured?

Several potential research issue suggestions arose regarding this topic: (a) Is transfer of training quantifiable? If so, how? If not, should it be used to compare the effectiveness of different devices? (b) Do training effectiveness measures tell us anything about learned cognitive, as opposed to psychophysical, behaviors? (c) Are laboratory data reliable in this context, or should the emphasis be placed on field data? (d) Should naive or experience subjects be used in experiments designed to measure transfer of training and training device effectiveness.

- How might mental models be used in designing training equipment?

Research is in progress on the use of mental models and cognitive mapping in flight simulator development. Given that results appear promising, it was suggested that efforts should be put into extending this line of research.

- Should research in training device development focus more on the issue of motivation than it has heretofore?

This issue relates to that of user acceptance. It appears that very little research on training devices has focused on motivation issues. An increased orientation towards analysis of factors affecting learner motivation was suggested.

- How should one interject new simulation ideas into an ongoing curriculum?

It was suggested that research is needed to determine ways to introduce a new training device or idea into an established curriculum or program which has been running smoothly, with a minimum of threat and disruption.

- Conduct research on development of a decision-making/selection model for use in designing training devices.

Instructional features, level of fidelity, etc., of a device are optimally considered prior to substantial investment costs in hardware development. It was suggested that if generalizable principles about device types could be established which would guide such decision, it may be possible to minimize the amount of research required for individual devices. Examples of the types of models suggested by this research issue are the media selection models used in ISD-type applications.

Criteria for Selecting among Research Issues

Given the constraints of time and money in training-oriented research programs, it is necessary to address the issue of criteria which might be applied to candidate research issues in order to prioritize them in terms of potential for funding.

The following criteria were suggested for use in prioritizing research. While not exhaustive, they appear to reflect major considerations which could be applied to research issue prioritization.

- o Cost of the research in terms of dollars, time, manpower and other resources
- o Likelihood of effective payoff
- o Feasibility of conducting the research
- o Sponsor priorities (i.e., policy)
- o Generalizability of the outcome
- o User acceptance
- o Training effectiveness potential
- o Technological gap closure

- o "Public relations" value/visibility
- o Operational requirements
- o Life Cycle cost of research product
- o Availability of other necessary resources

Research Support

The final issue addressed in the Priorities and Support session was that of generating the support necessary to conduct the long-term research required on training/simulation topics. The area of support was also addressed in terms of two general categories:

1. How can more dollars for research of this type (i.e., complex, long-range) be generated?
2. How can the flow and dissemination of information across the Services and disciplines be facilitated?

In terms of generating research dollars, three areas were emphasized for consideration. First, it was suggested that by educating foreign governments in the importance of research on training and simulation topics, additional cost might be added to the purchase price of U.S. military equipment, which monies would go directly toward the funding of such research. This could, however, increase the price of U.S. equipment to foreign governments and hence decrease its saleability. A second suggested source of potential research dollars lies within U.S. industry. If industry could be encouraged to invest a portion of their internal research and development (IR&D) money into research of this type, the support base could be greatly expanded. Finally, it was suggested that the limited research monies already available could perhaps, be more efficiently spent if there were a shift in emphasis from situation-specific simulators towards more generic simulators. Such a move could reduce the redundancy of existing systems and thus result in a net savings of monies already allocated for these purposes.

A second perspective on the issue of research support addressed support in terms of information dissemination. It was suggested that the more extensive dissemination of research data and related information, the lower the probability of future research being redundant. Several methods of maximizing the dissemination of information across disciplines and services were suggested:

- o Establishment of an analog to the Department of Agriculture's Field Service Representatives in the training/simulator area. Such individuals are needed in

both the utilization and design phases.

- o Increasing the availability of relevant literature by:
 - removing the limitations on publishing government studies;
 - establishing policy changes on dissemination of government documents;
 - establishing a centralized database on training and simulation issues.
- o Development of better guidebooks on training device development and transfer of training issues.
- o Educating the user community to the benefits of research on training issues through the open literature.

RESEARCH ON THE FIDELITY OF SIMULATORS

Jesse Orlansky
Institute for Defense Analyses
Arlington, Virginia 22202

PREVIOUS PAGE BLANK-NOT FILMED

ABSTRACT

The fidelity of a simulator is one of many features that may influence its effectiveness as a training device. The effectiveness of a particular amount or type of fidelity in a simulator should be measured by performance on the job, rather than primarily at school. The major issue in fidelity is one that involves a comparison of what a particular degree of fidelity may cost compared to what it may provide in improved training and job performance or in what it may save in the costs of training and job performance.

PURPOSE

The purpose of this workshop is to establish the amount of fidelity needed to make simulators effective for training. Since we do not know how to do that at present, we could start by trying to identify the research and development that is needed to help us approach that goal. Our work will be relevant and useful as long as we remember that the real issue is "How much fidelity do we need in simulators and training devices". Thus, we should design our experiments to tell us how much of the variance in training effectiveness is accounted for by the fidelity of the simulators used in various types of training.

Even though we focus on fidelity at this meeting, we know that many other factors can also influence the effectiveness of a simulator. These would include, for example, the relevance of the course content to the jobs the students are being trained to perform, the quality of the students and instructors, the fraction of total course time devoted to training with the simulator, and so on. The fidelity of a simulator may (or may not) be an important contributor to the effectiveness of a course. The relative importance of simulator fidelity will probably vary, thus, with how much of the course depends on it, what type of simulator is involved, and so on. I guess fidelity is important but it isn't everything.

Transfer of Training

The next issue concerns how to measure the effectiveness of simulators of different degrees of fidelity. (In the discussion that follows, it should be clear that we need a taxonomy to characterize different types of simulators and one or more scales with which to estimate amount of fidelity. It should also be clear that both issues are disregarded in the present discussion.) It is generally convenient to compare the effectiveness of two training devices used at a school in such terms as student grades or scores on tests, ratings by instructors, time needed to complete a course, and amount of student attrition. Such data are useful in deciding whether a student meets the standards established to qualify him for the job for which he is being trained; such data are also useful for various purposes related to managing a school. However, the real issue is not how well it prepares him to perform some real job. Transfer of training is simply a way of measuring how much of what is taught at school carries over to effective performance on the job. There are cases where the amount of effectiveness on the job, when related to what was carried over from a simulator used at school, varied from almost zero to 100 percent.

This means that we must measure the effect of simulator fidelity not on performance at school but on performance on the job. The case of platform motion in flight simulators offers an interesting and extremely relevant example. It turns out that

pilot performance in simulators improved together with increases in the fidelity of platform motion. The improvements in platform motion were due to the use of more accurate aerodynamic equations, better servo mechanisms, and increases in the number of degrees of motion being simulated (direction and rotation). This would argue that increased fidelity improved the effectiveness of the flight simulator.

However, a series of studies conducted since 1974 shows that pilot performance in the air is about the same for pilots trained in a simulator without motion compared to other pilots trained in the same simulator with motion (Koonce 1974, Jacobs and Roscoe 1975, Woodruff and Smith 1974, Gray and Fuller 1977, Woodruff et al 1976, and Martin and Waag 1978). This would argue that differences in the fidelity of flight simulators, with respect to platform motion only, do not influence its effectiveness for training provided, of course, we use a relevant measure of effectiveness. Performance in an aircraft is more relevant than performance in a flight simulator.

This example of what to measure should be an important guide to us in our work on fidelity. It is only fair to say that the results reported here are accepted generally for simulators of fighter aircraft, but it is still an open issue as to whether platform motion is needed for simulators of multi-engine aircraft.

Almost all military training, relating both to operations and maintenance, is based on the premise that on-the-job training supplements and increases the skills learned at school. Therefore, when devices of different levels of fidelity are compared, it is important to know (1) whether they produce different levels of performance at school, and (2) if such differences are found, how long they persist on the job. These possibilities are illustrated in the accompanying figures.

Figure 1 illustrates the case where two devices of different levels of fidelity produce about the same level of performance by the student at school and on the job. In this case, the less expensive device is as effective as the more expensive one. We have drawn a special case where performance at school was about the same, but where differences were observed later on the job; depending on how long these differences persisted, there might be a real difference in the effectiveness of the two assumed devices.

Figure 2 illustrates the case where different levels of fidelity produce a real difference at school and for some time on the job. Here, the issue is whether the cost of the amount of time it takes to produce equal performance on the job is more or less than the difference in cost of the two devices used at school.

As I understand the literature, studies in a number of areas of training have shown that significant differences in the fidelity of otherwise comparable simulators produce little

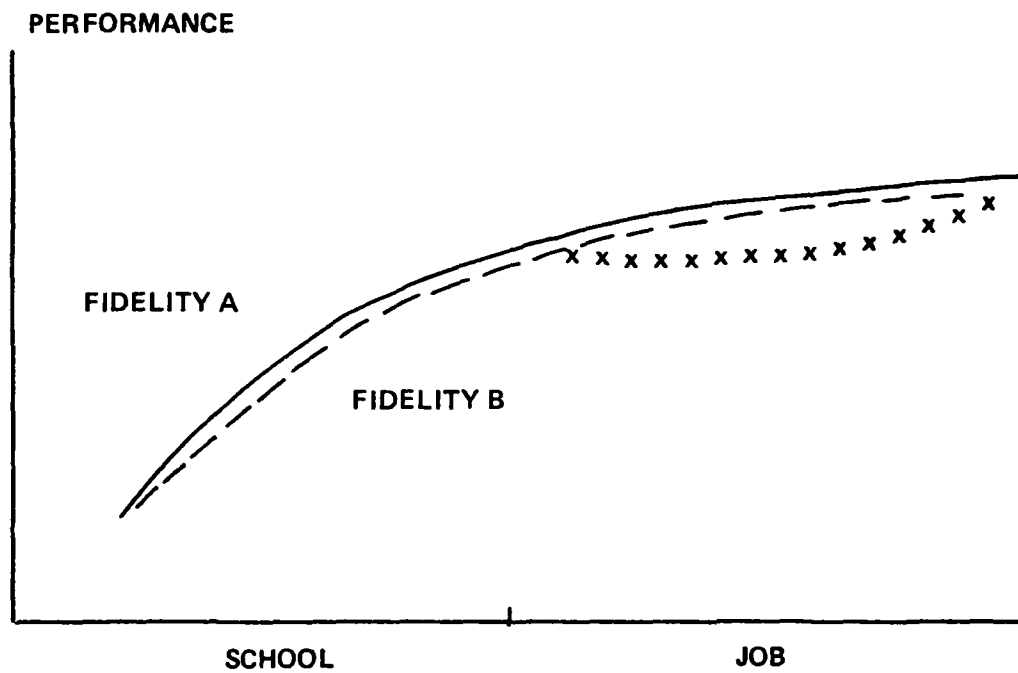


FIGURE 1
PERFORMANCE AT SCHOOL AND ON THE JOB WITH TRAINING DEVICES
OF TWO LEVELS OF FIDELITY (CASE 1).

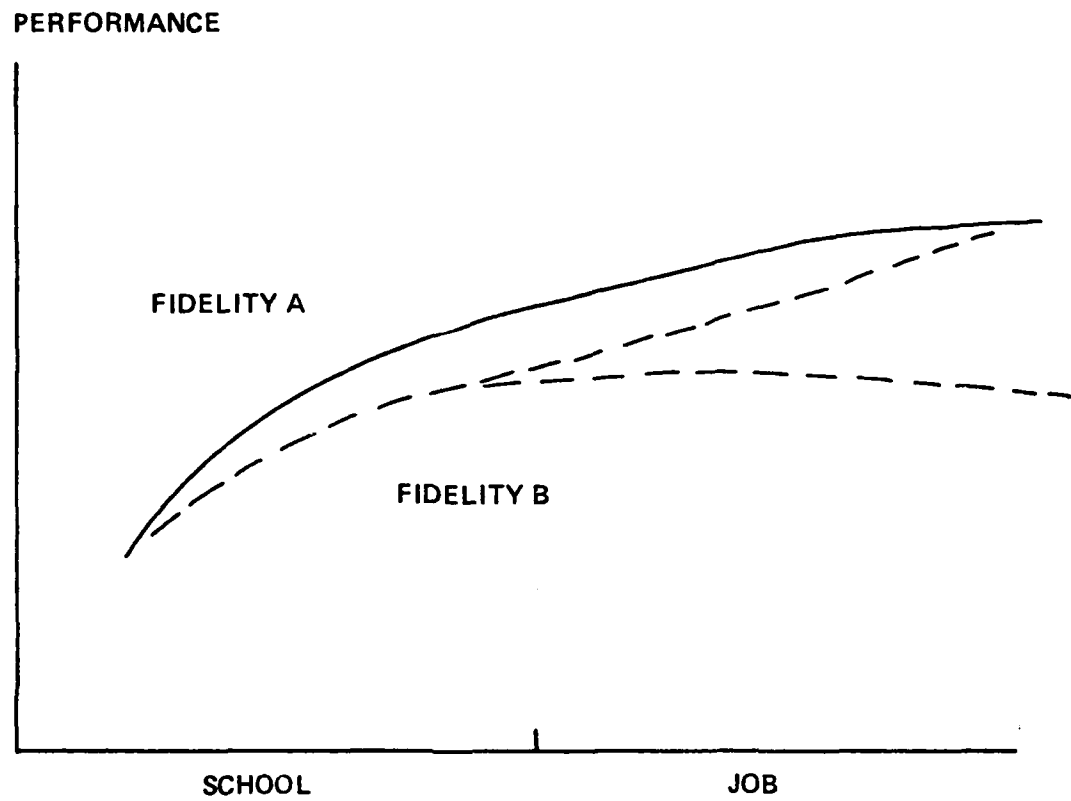


FIGURE 2
PERFORMANCE AT SCHOOL AND ON THE JOB WITH TRAINING DEVICES
OF TWO LEVELS OF FIDELITY (CASE 2)

difference in the effectiveness of training at school. Some studies have shown that differences in training procedures but not in fidelity of simulation produce significant differences in initial performance on the job; however, on-the-job experience helps the inferior performers to catch up quickly. That is, the superiority in performance is of relatively brief duration.

Measures of Effectiveness

There are many, many ways of measuring effectiveness, and this is not the place to discuss their reliability and relevance (i.e., validity). Since there is no general set of measures of effectiveness, it is likely that many different ones will be used in a variety of experiments. The practical issue is that it will be helpful to design our evaluations of fidelity so that they use multiple, overlapping measures of effectiveness. It will then, at least, be possible to compare the results of different studies and begin to develop some generalizations about the fidelity of simulators.

Costs

Although fidelity was the major topic discussed at the workshop, the underlying issue is one of costs. It is widely recognized that differences in the fidelity of simulators are really differences in the costs associated with building simulators of different degrees of engineering complexity. This means that, whether we like it or not, we must improve our understanding of how to compile and evaluate cost data that are relevant to fidelity of simulation. We will have to do this at least twice in every evaluation, i.e., once for each of the simulators with different degrees of fidelity whose effectiveness will be compared. The issues are far from trivial; we need to know the "cost-drivers", i.e., the cost of those components that most affect the fidelity of simulation and their relative cost compared to that of other components. We must also become interested in the costs of development versus the costs of use. The total life-cycle costs (that is, all the costs of development and use) and how long it takes to amortize the costs of new or improved simulators; it may be that some improvements in effectiveness due to improved fidelity may be so small that they are not worth what they cost to develop, acquire, and use.

Initial R&D

The issue of fidelity of simulation has been with us for over 25 years, at least since R.B. Miller (1954) wrote his paper on psychological considerations in the design of training equipment. The topic has been reviewed more than once, including Micheli (1972), Valverde (1973), and Semple et al (1981), among others. Most of these reviews have been concerned primarily with flight simulators although many other types of simulators are also used in training. There does not seem to be a great need for another literature review.

In general these studies reviewed the transfer of training of various types of simulators for various aspects of flight training. It might be helpful to re-compile these findings on the basis of the amount of transfer of training among devices that differed in their degree of fidelity. It would also be helpful to use the same index of transfer of training for all of these studies. It would be of interest to calculate several indices of transfer of training for each study and to correlate the results so as to improve our understanding of what the different indices of transfer measure. Finally, it would be helpful to perform a similar service for devices of differing degrees of fidelity in types of training that do not use flight simulators, e.g., maintenance, electronics, and vehicles.

This type of evaluation would be an empirical and productive way of initiating a research program in fidelity of simulation. It would avoid polemics. It would let the data tell us what we know and don't know and thus provide a straightforward way of shaping the steps that we should take now.

REFERENCES

- Gray, T.H. and R.R. Fuller; Effects of Simulator Training and Platform Motion on Air-to-Surface Weapons Delivery Training, Technical Report: AFHRL-TR-77-29, Air Force Human Resources Laboratory, Brooks Air Force Base, Texas, 1977.
- Jacobs, R.S. and S.N. Roscoe, Simulator Cockpit Motion and the Transfer of Initial Flight Training in New Concepts for Training Systems, 8th NTEC/Industry Conference Proceedings, Naval Training Equipment Center, Orlando, Florida, p. 113-122, 18-20 November 1975.
- Koonce, J.M. Effects of Gound-Based Aircraft Simulator Motion Conditions Upon Predictions of Pilot Proficiency, AFOSR-74-1292, University of Illinois at Urbana-Champaign, Aviation Research Laboratory, Savoy, Illinois, April 1974.
- Martin, E.L., and W.L. Waag, Contributions of Platform Motion to Training Effectiveness: Fidelity of Simulation of Training Equipment, TAEG Report 2, Training Equipment and Evaluation Group, Naval Training and Equipment Center, Orlando, Florida, 1972.
- Miller, R.B., Psychological Considerations in the Design of Training Equipment, WADC Technical Report 54-563, Wright Aeronautical Development Center, Ohio, 1954.
- Semple, C.A., Hennessy, R.T., Sanders, M.S., Cross, B.K., Beith, B.H., and McCauley, M.E. Aircrew Training Devices: Fidelity Features, Technical Report AFHRL-TR-80-36, Air Force Human Resources Laboratory, Brooks Air Force Base, Texas, January 1981.
- Valverde, H.H. A Review of Flight Simulator Transfer of Training Studies, Human Factors, 15, 510-523, 1973.
- Woodruff, R.R., and Smith J.F. T-4G Simulator and T-4 Ground Training Device in USAF Undergraduate Pilot Training, Technical Report AFHRL-TR-74-78, Air Force Human Resources Laboratory, Brooks Air Force Base, Texas, November 1974.
- Woodruff, R.R., Smith, J.F., Fuller, J.R. and Weyer, D.C. Full Mission Simulation in Undergraduate Pilot Training: An Exploratory Study, Technical Report AFHRL-TR-76-84, Air Force Human Resources Laboratory, Brooks AFB, Texas, December 1976.

SUMMARY REMARKS

David R. Baum
Honeywell Systems and Research Center
Minneapolis, Minnesota 55413

PRECEDING PAGE BLANK-NOT FILMED

ABSTRACT

The lack of a universally accepted paradigm for accomplishing training device design is identified as the major source of disagreement among participants at the workshop. A more integrated, systematically developed data base is needed to make progress towards a science of training device design. Several important points emerging from the workshop are briefly highlighted.

I want to thank you all for participating. I want to thank you for your patience and the contributions that each of you have made to the program that we have embarked on. I'd like to thank the topic leaders for their extremely acute and pertinent summaries. That was a very difficult task, and I'm very glad that we found four volunteers to take on that job. I'd like to thank you, Dr. Orlansky, for your summary remarks. I think that we need to take your concerns to heart and need to raise them to a higher level of attention than they sometimes get. We sometimes do lose sight of the forest for searching out the trees.

WHY IS THE PROBLEM COMPLEX?

Yesterday morning I determined that at least a portion of the audience was awake when I misspoke myself and said that we were taking a "bottoms down" approach to try to solve the problem. In retrospect that makes me feel comfortable because it certainly bounded the problem. And we've tried to deal with a very complex problem, a very complex set of issues. There was certainly some agreement on a number of key topics and a number of key issues in terms of what should receive emphasis in the Behavioral Science R&D community. Also there was agreement in what we have to offer ultimately to the folks who use our results to make decisions and to accomplish training.

I would submit that the type of discussions that we've had and disagreements that have emerged, the lack of a consensus that doesn't always get publicly stated, are an indication that training device design and development is still much more of an art involving common sense than it is based on well-established theory. There is at present no conceptual framework, there is no universe of discourse for defining training simulator development that we all ascribe to, that we all feel comfortable with. I don't intend this as a criticism, merely an observation. You don't start out in an area as complex as this with an accepted theory; that is something we must develop. We're pre-paradigmatic in the Kuhnian sense (Kuhn, 1962) when it comes to having procedures and techniques that are well founded empirically and that can be used to assure the people who have to accomplish training that they are going to be provided with effective training devices. In short, we lack a science of training device design.

I'd be willing to bet that taking any random subset of individuals here, they would be willing to guarantee that they could design a solution to a training problem for X number of dollars that would, if implemented, repay that number of dollars in some reasonable amount of time. That is the kind of guarantee we talked about in other areas in terms of guaranteeing a certain amount of transfer effectiveness and in terms of guaranteeing other aspects of the consequences of training. However, because our techniques and procedures tend to be somewhat idiosyncratic, each

solution would be different. Furthermore, assuming we had adequate measurements, one solution would prove superior. We need a better data base on which techniques and procedures yield superior training device solutions.

SOME KEY POINTS

What I want to do at this point is highlight several of the ideas expressed during the workshop that I think are important. It is just a listing and not an attempt to provide you with a conceptual framework or universe of discourse.

The notion of developing a family of trainers in the context of an overall training program, where any particular trainer is designed to accomplish training for a particular stage of learning or a particular stage of task acquisition, is, I think, really fundamental. We often, in the device acquisition process, get lost in the procurement of a particular device, losing sight of where it fits in the training program and losing sight of the fact that when it arrives on the scene, it's going to impact everything else in that training program. That to me, in terms of providing guidance to PM TRADE, in this instance, the Army's training device developer, that to me is a preeminent point that has to be made.

Jesse (Orlansky), your comments notwithstanding, fidelity seems to be something of a red herring. The notion of working backwards from what it will take to provide effective transfer of training seems to me to be a more promising approach. In order to do that, we have to know what the job is; if we know what the job is, if we have good techniques for analyzing the job into its component tasks and skills and knowledges, then there seems to be an assumption that we can, using accepted principles of learning, design a device that will ultimately transfer to that job. The techniques for doing that task description and analysis are part of what I meant when I talked about an art and the use of common sense. They are not that well developed, they tend to be somewhat decentralized in the way that they are applied and we need to work towards a more integrated and systematic approach to starting with what the job entails and feeding that back into the training device development process.

There are a number of other points that are of importance, for example, in the area of providing guidance, education and the development of a base of expertise. However, I don't think that I want to enumerate additional topics that are in need of research. They will certainly be incorporated into the program of research that we develop.

Again, I just want to thank you all for participating. I hope that you found it to be a somewhat educational experience; certainly I did. I think that, as these things go, because we're in a developing state in terms of a science of training device

design, we can expect to perhaps be together again in 5 to 10 years to review where we have been and assess where we need to go. I wouldn't be surprised at that. I don't think that this prospect should be viewed negatively. I think that it's just a consequence of dealing with a very complex set of problems and trying to bring some order out of what at times appears to be a chaotic situation.

REFERENCES

Kuhn, T.S., The Structure of Scientific Revolutions. Chicago:
The University of Chicago Press, 1962.

THE ARI FIDELITY WORKSHOP:
SUMMARY COMMENTS

Robert T. Hays
U.S. Army Research Institute

ABSTRACT

A summary of the workshop, as a whole, is presented. No attempt is made to cover all issues raised at the workshop. Rather issues are highlighted in three general areas: Approaches to Fidelity, Communications Issues and Research Issues. Conclusions are drawn about the workshop itself as well as about the conduct of future research.

The ARI workshop on "Research Issues in the Determination of Simulator Fidelity" brought a diverse group of individuals together to discuss a wide variety of topics (see organization and goals paper). Rather than reiterate these topics, the goal of this summary is to highlight several issues that emerged during the workshop, during the review of workshop transcripts and from subsequent discussions with workshop participants. It is not assumed that these are the only or even the most important issues. It is assumed, however, that they are important and should be documented, if only for heuristic purposes.

In this summary paper, issues are discussed under three headings. The first groups of issues deal with approaches to fidelity. Even a cursory examination of the papers in these proceedings will demonstrate that different individuals approach the issue of simulator fidelity with different meanings and goals. Discussions of the various approaches to fidelity were an important part of the workshop. The second group of issues deal with communication between different "players" in the ISD process. It is often the case that different individuals, who are concerned with their own phase of the ISD process, lack sufficient contact with their counterparts working on other ISD phases. The development of any instructional system can be facilitated if all persons can communicate with one another and take a more holistic approach to their individual functions. Finally, the third group of issues deal with areas thought to require future research and how the results of these research efforts should be incorporated into the ISD process.

Approaches To Fidelity

Workshop participants engaged in lively discussions throughout the workshop about the meaning of the term fidelity and also about how the concept should be used. These differences in approaches to fidelity may be illustrated as in Figure 1. This figure shows a two-factor description of the concept of simulator fidelity. It is the impression of this author that most individuals at the workshop fell somewhere on one of these continua.

The vertical axis of Figure 1 illustrates two diverse opinions concerning the meaning of fidelity. At one extreme, fidelity is narrowly demarcated. In this view, fidelity refers only to the physical and functional similarity of the training device to the actual equipment. At the other end of this continuum is the belief that there should be multiple types of fidelity, such as task fidelity, psychological fidelity, equipment fidelity, perceptual fidelity, cognitive fidelity, etc. In the opinion of this author, we can best proceed by dealing with fidelity as the physical and functional aspects of the training device and treat other "types" of fidelity as classes of variables which interact with these aspects of the device hardware. There is, at this time, no resolution of this issue.

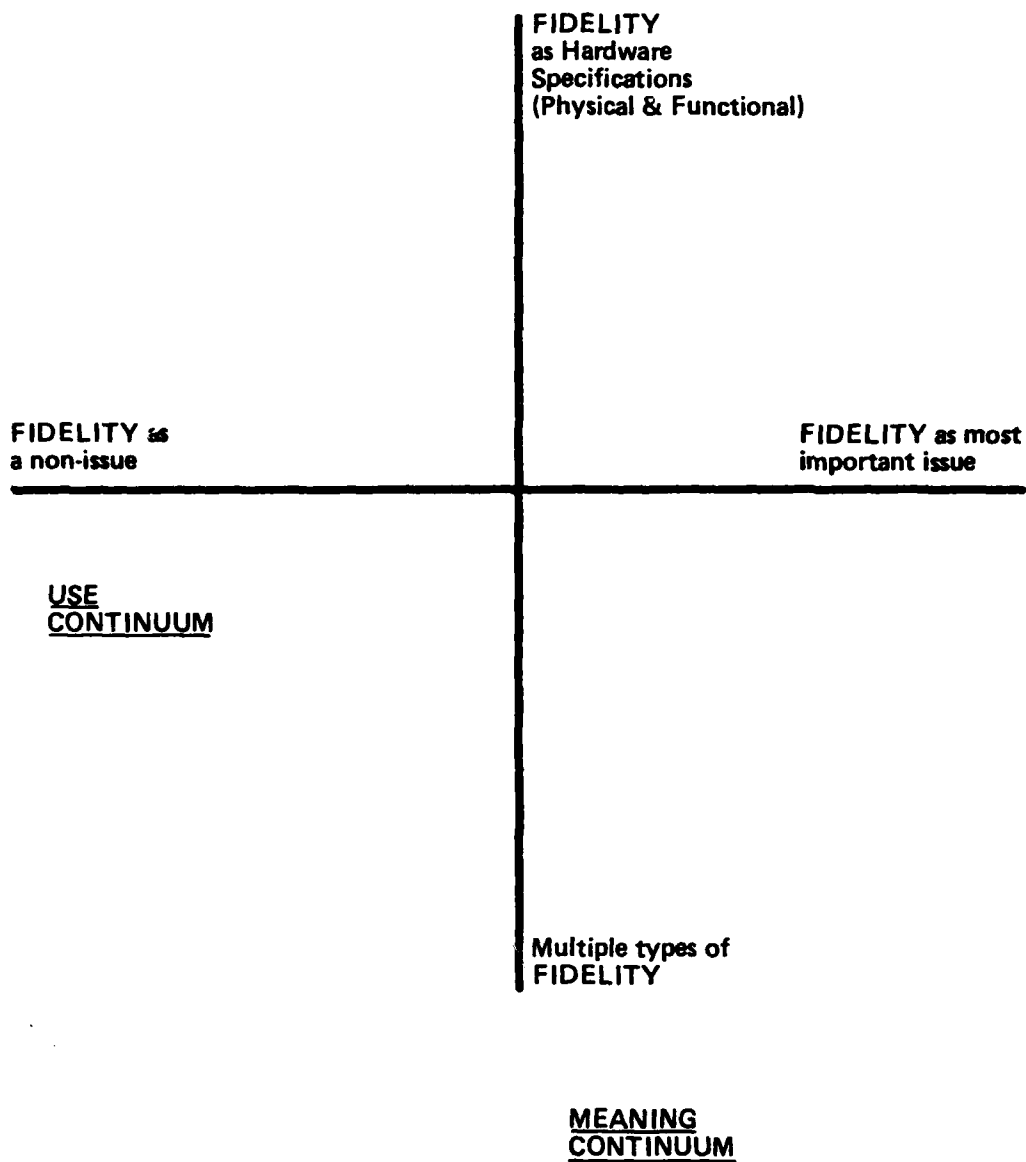


FIGURE 1
CONTINUUA ILLUSTRATING WORKSHOP PARTICIPANT'S APPROACHES
TO THE CONCEPT OF SIMULATOR FIDELITY

The horizontal axis of Figure 1 illustrates two diverse viewpoints over how the fidelity concept should be used in the design of training devices. One view is that the question of fidelity is very important since fidelity decisions are major contributors to the cost of training device development. Essentially this view states that without data demonstrating the value of low fidelity devices, large sums of money will be spent on high fidelity devices. The view represented by the other end of the horizontal continuum in Figure 1 is that fidelity is a non-issue. The real issue, according to this view, is to insure that the most effective instructional features are incorporated into any training device. This view eschews the use of the fidelity concept because it may force training device designers into equipment configurations that do not effectively incorporate the best instructional features.

The division between these views on the use of the fidelity concept may be semantic. Though one view does not use the term fidelity, the end product, after all instructional features are incorporated into the training device, is still a piece of training equipment with some degree of fidelity to the operational equipment which it simulates. Likewise, the training equipment that is designed to a given level of fidelity still must incorporate some form of instructional features.

These viewpoints can be reconciled if fidelity is treated as a summary concept on an ordinal scale. As such, it can provide a general overview of the entire training device which may be used during all phases of training device development. It can not be assumed that a fidelity specification will provide all information about the device. It is still necessary to develop, analyze and incorporate the best possible instructional features into the device. However, without becoming locked into a given degree of fidelity, the concept can be used, during front-end analyses, to provide designers with general specifications about what is believed to be necessary in the training device. The best available data may then be used to specify the details of the instructional features and other aspects of hardware and software for the training device. Once all the elements have been incorporated into the training device, fidelity may again be used to describe the configuration of the device as a whole for inter-device comparisons.

In the opinion of this author, the most important thing to remember about the concept of fidelity is to use it as a tool, but not to let it dictate to the user how it should be used. Fidelity does not necessarily imply high (or low) fidelity. It only has meaning as a summary concept, but as such, can still be used to advantage by instructional system developers.

Communication Issues

There are many phases of ISD that are not concerned directly with fidelity. However, it was quickly acknowledged by most

workshop participants that no matter how a device is designed, no matter what its level of fidelity, it will not be an effective trainer if it is not used properly. Likewise, it is not possible to design an effective training device if the task to be trained is not understood and if the context of instruction is not compatible with the training device.

Individuals who play different roles in ISD need to be educated on the limitations as well as the capabilities of any training device. The results of research on device configuration (i.e., fidelity) must be communicated to curriculum developers, engineers, instructors and all the other persons involved in ISD in a form that can be used in the decision processes that each must make. It is also necessary to communicate the needs of the training community to researchers so they may design relevant experiments.

An idea, which may be of value in improving the public relations of training device acceptance and providing important information about training device design and use, was raised during the workshop. The ISD process might include a role analogous to the Department of Agriculture extension agent. This individual would serve as a training device facilitator, and could provide input to persons at any stage of device development or use. Information could be supplied to instructors about instructional features of a device, to curriculum developers about how to best incorporate the device into a POI, to engineers in designing and constructing the device, to researchers about what questions in device design to explore, as well as to the individuals who purchase the device. Most important, this facilitator would follow a device through all ISD stages and be able to provide a vital communications link between individuals who may only be concerned with narrow problems, but who could operate more effectively with a better overview of the whole ISD process.

Research Issues

Many divergent viewpoints were raised concerning how research should be conducted on the design of training devices. Some workshop participants felt that research should be conducted on how mental models about the device and about what the device is training, are instilled in trainers and instructors. Others expressed the need for research on how the devices are used in the schools to determine if instructional features are being utilized. Still others felt that research should be conducted to determine just which tasks should be trained with simulators and which should be trained with other devices or with actual equipment.

Most participants were in agreement that empirical data are both necessary and lacking. Many felt that these data should be generated in transfer of training (TOT) experiments, but several participants believed that the problems with TOT designs outweigh their value. At this stage, various government, academic and

private organizations are attempting to provide unifying themes or structures for simulators and training device research. For example, ARI and PM TRADE are currently developing a joint long range program to integrate basic and applied research with PM TRADE's training device acquisition activities. In addition to this effort, AFHRL is evolving an information system on human performance designed to help equipment developers (including training device developers) make design decisions. The general belief is that until such structures are developed, a great deal of progress in providing guidance for training device development will not be made.

Several desirable characteristics for unifying structures for empirical research emerged from workshop discussions. They should at least: (1) Take a whole system approach providing for not only front-end analysis, but also in-progress analysis and end-of-development analysis (How the device is implemented). (2) Include basic research on the mental models of both trainees and instructors in terms of both how these models are acquired and how a given model affects training device usage. (3) Include research on how various device configurations (i.e., levels of fidelity) interact with a large variety of variables, such as task difficulty or level of device acceptance, to produce the desired level of transfer of training. (4) Be organized so that the data are in a form that is usable by diverse groups. (5) Be organized so that later efforts can build on previous research, constantly improving the empirical data base upon which training device design and implementation decisions are made.

Conclusion

The most obvious conclusion that may be drawn from this workshop is that there are many questions about training simulator fidelity that remain to be answered. The enthusiasm which workshop participants brought to these sessions provides hope that answers to a substantial number of these questions may be forthcoming. If the research community, the training community, the training device development community as well as the community which procures training devices can continue to communicate through workshops such as this one, a unified research paradigm may be developed which will support the needs of all concerned. This research paradigm must include a systematic approach which incorporates existing data as well as generating new data on all of the complex interactions between device configurations (fidelity) and transfer of training. The ultimate goal of this research will be to provide guidance for device developers in the form of prescriptive and predictive models which can be used to design and evaluate training devices at all stages of the Life Cycle System Management Model (LCSMM) so that the effectiveness of any training system can be maximized.

LIST OF ATTENDEES
ARI FIDELITY WORKSHOP
23-24 July 1981

Alluisi, Earl; AFHRL/CCN, AV 240-3605
Barber, Herbert; Ft. Leavenworth, AV 552-5689
Baum, Dr. David R.; Honeywell SRC, 612-378-5046
Bessemer, David; ARI-Knox, AV 464-4932
Bickley, Will; ARI, AV 558-2873
Boldovici, John; ARI, AV 284-8836
Boneau, C. Alan; George Mason Univ, 703-323-2203
Boycan, Gary; ARI, AV 284-9420
Brock, John F.; Honeywell SRC, 612-378-4403
Buckland, MAJ George; AFHRL/OT, AV 474-6561
Campbell, Frank; Dir. of Training Development,
Ft. Rucker, AV 558-7111
Campbell, Dr. J. Olin; WICAT, 801-224-6400
Caro, Dr. Paul W.; Seville Research Corporation
904-434-524
Chambers, Walter S.; NTEC, AV 791-5353
Chatelier, CDR Paul; OUSDR&E, 225-9777
Cicchinelli, Dr. Louis F.; Denver Research Inst/SSRE
303-753-2809
Clayton, MAJ L.W.; USAF/RDIT, AV 858-4433
Clement, Warren; Systems Technology Inc.; 415-961-4674
Connelly, Ed; Performance Measurement Assoc., Inc
Davis, Elaine; Science Applications, Inc. 703-734-5872
Evans, Bob; ARI, AV, 284-8836
Hagman, Dr. Joseph; ARI, AV, 284-8695
Harris, Dr. Frank; ARI, AV 284-8827

Hays, Dr. Robert; ARI, AV 284-8836

Hennessey, Dr. Robert T.; National Research Council;
202-334-3027

Hirshfeld, Dr. Stephen; Honeywell SRC; 612-378-5377

Holman, Dr. Garvin; ARI; AV 284-9420

Houston, Thomas; ARI; AV 284-8836

Hritz, Rohn J; Applied Science Assoc; 412-586-7771

Johnson, Dr. Richard; ARI; AV 284-8943

Kahn, Otto; ARI-Monterey; AV 929-8316

Kane, Dr. John; SAI; 703-827-4862

Kennedy, Dr. Bob; Canyon Research Group; 305-894-5090

Kimball, Gerry; Honeywell T&CSO; 213-331-0011; x 2818

Klein, Dr. Gary; Klein Assoc; 513-767-2691

Lattimore, Andy; US Army Trng Support Center; AV
927-3564

Laughery, Ronald K.; Calspan, 716-632-7500

Laurence, Dr. Michael T.; ARI; AV 284-8836

Legum, Stan; Kinton; 703-820-1000

Magee, L.E., DCIEM (CANADA); 416-633-4240, x 306

Martin, Dr. Ed; ASD/ENETS; AV 785-4408

McLean, Craig; ASD/YWB; 785-2885

McMichael, Dr. Jim; Navy Personnel Research & Development Center,
AV 933-7121

Mirabella, Dr. Angelo; ARI; AV 284-8827

Montague, Dr. William; Navy Personnel Research & Development
Center, AV 933-7121

Muller, William; Honeywell SRC; 612-378-5086

Nawrocki, Dr. Leon H.; ARI Scientific Coordination Office, AV
680-3742

Olmstead, Dr. Joseph; HumRRO; 703-549-3611
Orlansky, Dr. Jesse; Institute for Defense Analyses, 703-558-1660
Pasnak, Dr. Robert; George Mason University, 703-250-6227
Peckman, Don; PM TRADE, AV 791-5881
Rizzo, Bill; NTEC; AV 791-5130
Root, Robert; ARI, AV 284-8721
Rose, Andrew; AIR, 617-275-0800
Scanland, Dr. F. Worth; CNET (Code N-5); AV 933-3591
Schendel, Dr. Joel; ARI, Ft. Benning; AV 835-1360
Semple, C.A.; Canyon Research Group, Inc., 213-889-5072
Siecko, Nicholas A.; ECC; 215-687-2600
Smith, Deb; Honeywell; 612-378-5775
Smith, Ed; AFHRL, AV 370-3391
Swezey, Dr. Robert; SAI, 703-827-4974
Thorpe, MAJ Jack; Cybernetics Technology Div/DARPA, 202-694-1303
Tufano, Dr. Dan; ARI, AV 284-8836
Winter, David; ARI, 617-275-0800
Wood, LTC Piers; M., USATSC, AV 927-4681
Zacharias, Dr. Greg; Bolt, Beranek & Newman, Inc., 617-497-3397
Zeidner, Dr. Joseph; ARI, AV 284-8636